



Template Shape Estimation: Correcting an Asymptotic Bias

Nina Miolane, Susan Holmes, Xavier Pennec

► To cite this version:

Nina Miolane, Susan Holmes, Xavier Pennec. Template Shape Estimation: Correcting an Asymptotic Bias. SIAM Journal on Imaging Sciences, 2017, 10 (2), pp.808 - 844. 10.1137/16M1084493 . hal-01350508v2

HAL Id: hal-01350508

<https://hal.science/hal-01350508v2>

Submitted on 2 Feb 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Template shape estimation: correcting an asymptotic bias*

Nina Miolane[†], Susan Holmes[‡], and Xavier Pennec[†]

Abstract. We use tools from geometric statistics to analyze the usual estimation procedure of a template shape. This applies to shapes from landmarks, curves, surfaces, images etc. We demonstrate the asymptotic bias of the template shape estimation using the stratified geometry of the shape space. We give a Taylor expansion of the bias with respect to a parameter σ describing the measurement error on the data. We propose two bootstrap procedures that quantify the bias and correct it, if needed. They are applicable for any type of shape data. We give a rule of thumb to provide intuition on whether the bias has to be corrected. This exhibits the parameters that control the bias' magnitude. We illustrate our results on simulated and real shape data.

Key words. shape, template, quotient space, manifold

AMS subject classifications. 53A35, 18F15, 57N25

Introduction. The shape of a set of points, the shape of a signal, the shape of a surface, or the shapes in an image can be defined as the remainder after we have filtered out the position and the orientation of the object [24]. Statistics on shapes appear in many fields. Paleontologists combine shape analysis of monkey skulls with ecological and biogeographic data to understand how the *skull shapes* have changed in space and time during evolution [16]. Molecular Biologists study how *shapes of proteins* are related to their function. Statistics on misfolding of proteins is used to understand diseases, like Parkinson's disease [29]. Orthopaedic surgeons analyze *bones' shapes* for surgical pre-planning [11]. In Signal processing, the *shape of neural spike trains* correlates with arm movement [25]. In Computer Vision, classifying *shapes of handwritten digits* enables automatic reading of texts [5]. In Medical Imaging and more precisely in Neuroimaging, studying *brain shapes* as they appear in the MRIs facilitates discoveries on diseases, like Alzheimer [30].

What do these applications have in common? Position and orientation of the skulls, proteins, bones, neural spike trains, handwritten digits or brains do not matter for the studies' goal: only *shapes* matter. Mathematically, the study analyses the statistical distributions of *the equivalence classes of the data* under translations and rotations. They project the data in a quotient space, called the *shape space*.

The simplest - and most widely used - method for summarizing shapes is the computation of the mean shape. Almost all neuroimaging studies start with the computation of the mean brain shape [18] for example. One refers to the mean shape with different terms depending on the field: mean configuration, mean pattern, template, atlas, etc. The mean shape is an average of *equivalence classes of the data*: one computes the mean after projection of the data in the shape space. One may wonder if the projection biases the statistical procedure. This

*Accepted for publication in SIAM Journal of Imaging Science. Submitted to the editors on January 23, 2017. This work was funded by the virtual lab Inria@SiliconValley.

[†]INRIA, Asclepios project-team, 2004 Route des Lucioles, 06902 Sophia Antipolis, France, nina.miolane@inria.fr

[‡]Stanford University, Department of Statistics, Sequoia Hall, Serra Mall, Stanford CA

is a legitimate question as any bias introduced in this step would make the conclusions of the study less accurate. If the mean brain shape is biased, then neuroimaging’s inferences on brain diseases will be too. This paper shows that a bias is indeed introduced in the mean shape estimation under certain conditions.

Related work. We review papers on the shape space’s geometry as a quotient space, and existing results on the mean shape’s bias.

Shapes of landmarks: Kendall analyses The theory for shapes of *landmarks* was introduced by Kendall in the 1980’s [23]. He considered shapes of k labeled landmarks in \mathbb{R}^m . The size-and-shape space, written $S\Sigma_m^k$, takes also into account the overall size of the landmarks’ set. The shape space, written Σ_m^k , quotients by the size as well. Both $S\Sigma_m^k$ and Σ_m^k have a Riemannian geometry, whose metrics are given in [27]. These studies model the probability distribution of the data directly in the shape space Σ_m^k . They do not consider that the data are observed in the space of landmarks $(\mathbb{R}^m)^k$ and projected in the shape space Σ_m^k . The question of bias is not raised.

We emphasize the distinction between ”form” and ”shape”. ”Form” relates to the quotient of the object by rotations and translations only. ”Shape” denotes the quotient of the object by rotations, translations, and scalings. Kendall shape spaces refer to ”shape”: the scalings are quotiented by constraining the size of the landmarks’ set to be 1.

Shapes of landmarks: Procrustean analyses Procrustean analysis is related to Kendall shape spaces but it also considers shapes of landmarks [19, 12, 20]. Kendall analyses project the data in the shape space by explicitly computing their coordinates in Σ_m^k . In contrast, Procrustean analyses keep the coordinates in $(\mathbb{R}^m)^k$: they project the data in the shape space by ”aligning” or ”registering” them. Orthogonal Procrustes analysis ”aligns” the sets of landmarks by rotating each set to minimize the Euclidean distance to the other sets. Procrustean analysis considers the fact that the data are observed in the space $(\mathbb{R}^m)^k$ but does not consider the geometry of the shape space.

The mean ”shape” was shown to be consistent for shapes of landmarks in 2D and 3D in [28, 26]. Such studies have a generative model with a scaling component α and a size constraint in the mean ”shape” estimation procedure, which prevents the shapes from collapsing to 0 during registration. In contrast, the mean ”form” - i.e. without considering scalings - is shown to be inconsistent in [28] with an *reductio ad absurdum* proof. However, this proof does not give any geometric intuition about how to control or correct the phenomenon. More recently, similar inconsistency effects have been observed in [13], showing that implementing ordinary Procrustes analysis without taking into account noise on the landmarks may compromise inference. The authors propose a conditional scoring method for matching configurations in order to guarantee consistency.

Shapes of curves Curve data are projected in their shape space by an alignment step [22], in the spirit of a Procrustean analysis. The bias of the mean shape is discussed in the literature. Unbiasedness was shown for shapes of signals in [25] but under the simplifying assumption of no measurement error on the data. Some authors provide examples of bias when there is measurement error [2]. Their experiments show that the mean signal shape may converge to pure noise when the measurement error on simulated signals increases. The bias is proven in [7] for curves estimated from a finite number of points in the presence of error.

But again, no geometric intuition nor correction strategy is given.

Abstract shape spaces [21] studies statistics on abstract shape spaces: the shapes are defined as equivalence classes of objects in a manifold M under the isometric action of a Lie group G . This unifies the theory for shapes of landmarks, of curves and of surfaces described above. [21] introduces a generalization of Principal Component Analysis to such shape spaces and does not compute the mean shape as the 0-dimensional principal subspace. Therefore, the bias on the mean shape is not considered.

But in the same abstract setting, [33] shows the bias of the mean shape, in the special case of a finite-dimensional flat manifold M . The authors emphasize how the bias depends on the noise σ on the measured objects, more precisely on the ratio of σ with respect to the overall size of the objects. [4] also presents a case study for an infinite dimensional flat manifold M quotiented by translations, where the noise σ is one of the crucial variables controlling the bias. However, the case of general curved manifolds M has not been investigated yet.

Contributions and outline. We are still missing a *global geometric* understanding of the bias. Which variables control its magnitude? Is it restricted to the mean shape or does it appear for other statistical analyses? How important is it in practice: do we even need to correct it? If so, how can we correct it? Our paper addresses these questions. We use a geometric framework that unifies the cases of landmarks, curves, images etc.

Contributions. We make three contributions. First, we show that statistics on shapes are biased when the data are measured with error. We explicitly compute the bias in the case of the mean shape. Formulated in the Procrustean terminology, our result is: the Generalized Procrustes Analysis (GPA) estimator of mean "form" is asymptotically biased, because we do not consider scalings. Second, we offer an interpretation of the bias through the geometry of the shape space. In applications, this aids in deciding when the bias can be neglected in contrast with situations when it must be corrected. Third, we leverage our understanding to suggest several correction approaches.

Outline. The paper has four Sections. Section 1 introduces the geometric framework of shape spaces. Section 2 presents our first two contributions: the proof and geometric interpretation of the bias. Section 3 describes the procedures to correct the bias. Section 4 validates and illustrates our results on synthetic and real data.

1. Geometrization of template shape estimation.

1.1. Two running examples. We introduce two simple examples of shape spaces that we will use to provide intuition.

First, we consider two landmarks in the plane \mathbb{R}^2 (Figure 1 (a)). The landmarks are parameterized each with 2 coordinates. For simplicity we consider that one landmark is fixed at the origin on \mathbb{R}^2 . Thus the system is now parameterized by the 2 coordinates of the second landmark only, e.g. in polar coordinates (r, θ) . We are interested in the shape of the 2 landmarks, i.e. in their distance which is simply r .

Second, we consider two landmarks on the sphere S^2 (Figure 1 (b)). One of the landmark is fixed at the north pole of S^2 . The system is now parameterized by the 2 coordinates of the second landmark only, i.e. (θ, ϕ) , where θ is the latitude and ϕ the longitude. The shape of the two landmarks is the angle between them and is simply θ .

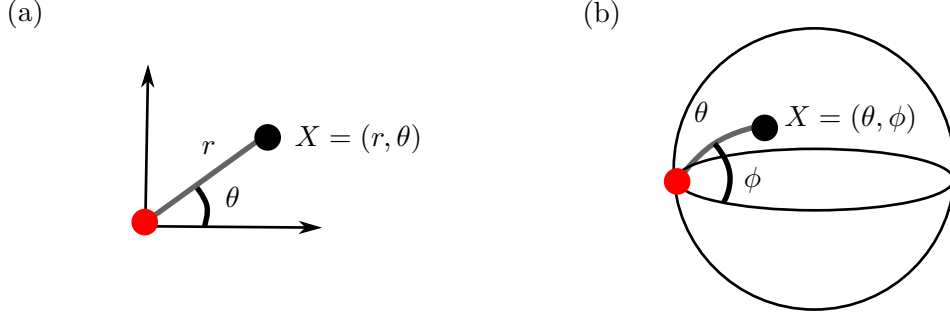


Figure 1. Two landmarks, one in red and one in black, on the plane \mathbb{R}^2 (a) and on the sphere S^2 (b). The landmark in red is fixed at the origin of the coordinates. The system is entirely represented by the coordinates X of the landmark in black.

1.2. Differential Geometry of shapes.

1.2.1. The shape space is a quotient space. The data are objects $\{X_i\}_{i=1}^n$ that are either sets of landmarks, curves, images, etc. We consider that each object X_i is a point in a Riemannian manifold M . In this paper, we restrict ourselves to finite dimensional manifolds. We have $M = \mathbb{R}^2$ in the plane example: a flat manifold of dimension 2. We have $M = S^2$ in the sphere example: a manifold of constant positive curvature and of dimension 2.

By definition, the objects' shapes are their equivalence classes $\{[X_i]\}_{i=1}^n$ under the action of some finite dimensional Lie group G : G is a group of continuous transformations that models what does not change the shape. The action of G on M will be written with \cdot . In our examples, the rotations are the transformations that leave the shape of the systems invariant. Let us take g a rotation. The action of g on the landmark X is illustrated by a blue arrow in Figures 2 (a) for the plane and (d) for the sphere. We observe that the action does not change the shape of the systems: the distance between the two landmarks is preserved in (a), the angle between the two landmarks is preserved in (d). The equivalence class of X_i is also called its orbit and written O_{X_i} . The equivalence class/orbit of X is illustrated with the blue dotted circle in Figure 2 (a) for the plane example and in Figure 2 (d) for the sphere example. The orbit of X in M is the submanifold of all objects in M that have the same shape as X . The curvature of the orbit as a submanifold of M is the key point of the results in Section 2.

The *shape space* is by definition the space of orbits. This is a quotient space denoted $Q = M/G$. One orbit in M , i.e. one circle in Figure 2 (b) or (e), corresponds to a point in Q . The shape space is $Q = \mathbb{R}_+$ in the plane example. This is the space of all possible distances between the two landmarks, see Figure 2 (c). The shape space is $Q = [0, \pi]$ in the sphere example. This is the space of all possible angles between the two landmarks, see Figure 2 (f).

1.2.2. The shape space is a metric space. We consider that the action of G on M is *isometric with respect to the Riemannian metric of M* . This implies that the distance d_M between two objects in M does not change if we transform both objects in the same manner. In the plane example, rotating the landmark X_1 and another landmark X_2 with the same angle does not change the distance between them.

The distance in M induces a quasi-distance d_Q in Q : $d_Q(O_{X_1}, O_{X_2}) = \inf_{g \in G} d_M(g \cdot X_1, X_2)$ [21]. The Lie group the action being isometric, the quasi-distance is in fact a distance. In the case of the The distance between the shapes of X_1 and X_2 is computed by first registering/aligning X_1 onto X_2 by the minimizing g , and then using the distance in the ambient space M . In the plane example, the distance between two shapes is the difference in distances between the landmarks. One can compute it by first aligning the landmarks, say on the first axis of \mathbb{R}^2 , then one uses the distance in \mathbb{R}^2 .

1.2.3. The shape space has a dense set of principal shapes. The *isotropy group* of X_i is the subgroup of transformations of G that leave X_i invariant. For the plane example, every $X_i \neq (0, 0)$ has isotropy group the identity and $(0, 0)$ has isotropy group the whole group of 2D rotations. Objects on the same orbit, i.e. objects that have the same shape, have conjugate isotropy groups.

Principal orbits or principal shapes are orbits or shapes with smallest isotropy group conjugation class. In the plane example, $\mathbb{R}^2 \setminus (0, 0)$ is the set of objects with principal shapes. Indeed, every X in $\mathbb{R}^2 \setminus (0, 0)$ belongs to a circle centered at $(0, 0)$ and has isotropy group the identity. The set of principal shapes corresponds to \mathbb{R}_+^* in the shape space and is colored in blue on Figure 2 (c). *Singular orbits or singular shapes* are orbits or shapes with larger isotropy group conjugation class. In the plane example, $(0, 0)$ is the only object with singular shape. It corresponds to 0 in \mathbb{R}_+ and is colored in red in Figure 2 (c).

Principal orbits form an open and dense subset of M , denoted M^* . This means that there are objects with non-degenerated shapes almost everywhere. In the plane example, $\mathbb{R}^2 \setminus (0, 0)$ is dense in \mathbb{R}^2 . In the sphere example, $S^2 \setminus \{(0, 0), (\pi, 0)\}$ is dense in S^2 , where $(0, 0)$ denotes the north pole and $(\pi, 0)$ the south pole of S^2 . Likewise, principal shapes form an open and dense subset in Q , denoted Q^* . In the plane example, \mathbb{R}_+^* is dense in \mathbb{R}_+ . In the sphere example, $]0, \pi[$ is dense in $[0, \pi]$.

The dense set M^* makes the projection in the quotient space a Riemannian submersion [21], which we use to embed $Q^* = M^*/G$ in M^* . In other words, the tangent space of the quotient space is identified everywhere with the vertical space with respect to the (isometric) Lie group action. Regular shapes of Q^* are embedded in the space of objects with regular shapes M^* . The computations in Section 2 will be carried out on the dense set M^* of principal orbits. The curvature of these principal orbits - i.e. of the blue circles of Figures 2(b) and (e) - will be the main geometric parameter responsible for the asymptotic bias studied in this paper. We note that the curvature of principal orbits is closely related to the presence of singular orbits: principal orbits wrap around the singular orbits. In the plane example, any blue circle - i.e. any principal orbit - wraps around its center, the red dot $(0, 0)$ - which is the singular orbit, see Figure 2(b).

We have focused on an intuitive introduction of the concepts. We refer to [38, 1, 21] for mathematical details. From now on, the mathematical setting is the following: we assume a proper, effective and isometric action of a finite dimensional Lie group G on a finite dimensional complete Riemannian manifold M .

1.3. Geometrization of generative models of shape data. We recall that the data are the $\{X_i\}_{i=1}^n$ that are sets of landmarks, curves, images, etc. In the general case, one can

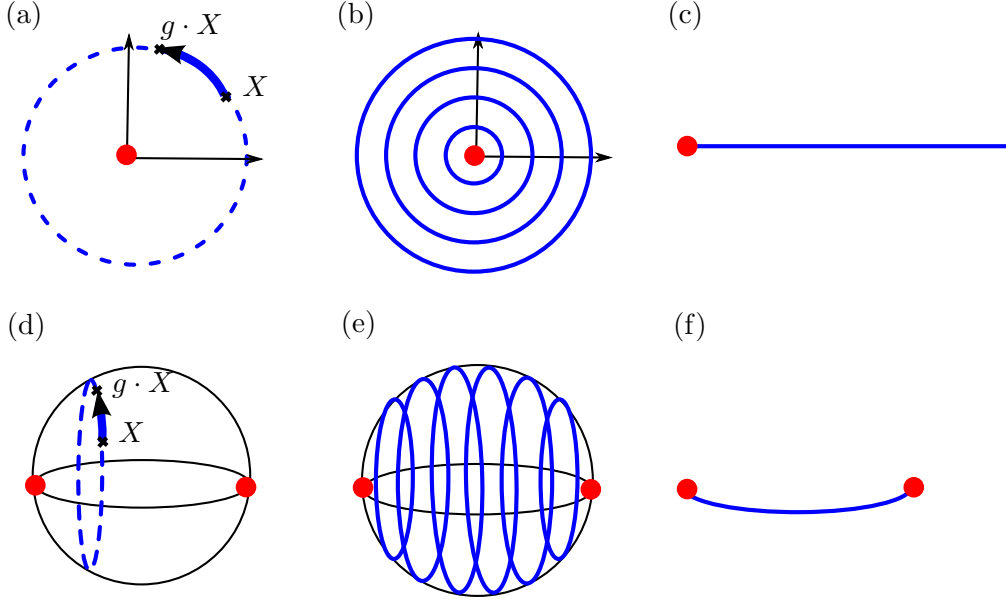


Figure 2. First line: Action of rotations on \mathbb{R}^2 , with (a): action of rotation $g \in SO(2)$ on point $X \in \mathbb{R}^2$ and orbit of X in blue dotted line; (b) Stratification of \mathbb{R}^2 into principal orbit type (blue) and singular orbit type (red); (c) shape space $\mathbb{R}_+ = \mathbb{R}^2/SO(2)$ with a singularity (red dot). Second line: Action of $SO(2)$ on S^2 with (d): action of rotation $g \in SO(2)$ on point $X \in S^2$ and orbit of X in blue dotted line; (e) Stratification of S^2 into principal orbit type (blue) and singular orbit type (red) (f) shape space $[0, \pi] = S^2/SO(2)$ with two singularities (red dots).

interpret the data X_i 's as random realizations of the generative model:

$$(1) \quad X_i = \text{Exp}(g_i \cdot Y_i, \epsilon_i) \quad i = 1 \dots n,$$

where $\text{Exp}(p, u)$ denotes the Riemannian exponential of u at point p . The Y_i , g_i , ϵ_i are respectively i.i.d. realizations of random variables that are drawn independently.

In this paper as well as often in the literature [2, 4, 8, 7, 25], we consider mainly the following simpler generative model:

$$(2) \quad X_i = \text{Exp}(g_i \cdot Y, \epsilon_i) \quad i = 1 \dots n,$$

where Y is a parameter which we call the template shape. The following three step formulation of the generative models 1 and (2) gives technical details and their interpretation in terms of shapes.

Step 1: Generate the shape $Y_i \in M^*/G \subset M^*$. In the full generative model (1), we assume that there is a probability density of shapes in the Riemannian manifold $Q^* = M^*/G$, with respect to the measure on Q^* induced by the Riemannian measure of M^* . The Y_i 's are i.i.d. samples drawn from this distribution. For example, it can be a Gaussian - or one of its generalization to manifolds [36] - as illustrated in Figure 3 on the shape spaces for the plane and sphere examples. This is the variability that is meaningful for the statistical study, whether we are analyzing shapes of skulls, proteins, bones, neural spike trains, handwritten digits or brains.

We mainly assume in this paper the simpler generative model (2) with parameter: the template shape $Y \in M^*/G$. In other words, we assume that the probability distribution is singular and more precisely that it is simply a Dirac at Y . This is the most common assumption within the model (1) [2, 3, 25, 8]. We point out that Y is a point of the shape space M^*/G , which is embedded in the object space by $M^*/G \subset M^*$, see previous subsection.



Figure 3. Step 1 of generative model of Equation (1) for the plane example (a) and the sphere example (b). The black curve illustrates the probability distribution function on shape space. This is a distribution on $r \in \mathbb{R}_+$ for the plane example (a) and on $\theta \in [0, \pi]$ for the sphere example. The black square represents its expectation. For the simpler generative model of Equation (2), the probability distribution boils down to a single point at Y i.e. at the black square.

Step 2: Generate its position/parameterization $g_i \in G$, to get $g_i \cdot Y \in M^*$. We cannot observe shapes in $Q = M/G$. We rather observe objects in M , that are shapes posed or parameterized in a certain way. We assume that there is a probability distribution on the positions or parameterizations of G , or equivalently a probability distribution on principal orbits with respect to their intrinsic measure. We assume that the distribution does not depend on the shape Y_i that has been drawn. The g_i 's are i.i.d. from this distribution. For example, it can be a Gaussian - or one of its generalization to manifolds [36] - as illustrated in Figure 4 on the shape spaces for the plane and sphere examples.

The drawn g_i is used to pose/parameterize the shape Y_i drawn in Step 1 (in the case of model of Equation (1)), where $Y_i = Y$ (in the case of model of Equation (2)). The shape is posed/parameterized through the isometric action of G on $Q^* \subset M^*$, to get the object $g_i \cdot Y_i \in M^*$, or the object $g_i \cdot Y \in M^*$ in the case of the simpler model of Equation (2).

Step 3: Generate the noise $\epsilon_i \in T_{g_i \cdot Y_i} M$. The observed X_i 's are results of noisy measurements. We assume that there is a probability distribution function on $T_{g_i \cdot Y_i} M$ representing the noise. We further assume that this is a Gaussian - or one of its generalization to manifolds [36] - centered at $g_i \cdot Y_i$, the origin of the tangent space $T_{g_i \cdot Y_i} M$, and with standard deviation σ , see Figure 5. The parameter σ will be extremely important in the developments of Section 2, as we will compute Taylor expansions around $\sigma = 0$.

Other generative models may be considered in the literature. We find in [3] the model: $X_i = g_i \cdot \text{Exp}(Y_i, \epsilon_i)$, where the Riemannian exponential Exp is also performed in M through the embedding $Y_i \in Q \subset M$. In [25], we find the model without noise: $X_i = g_i \cdot Y_i$.

1.4. Learning the variability in shapes: estimating the template shape. Our goal is to unveil the variability of shapes in $Q = M/G$ while we in fact observe the noisy objects X_i 's in M . We focus on the case where the variability in the shape space is assumed to be a Dirac at Y . Our goal is thus to estimate the template shape Y , which is a parameter of the generative model.

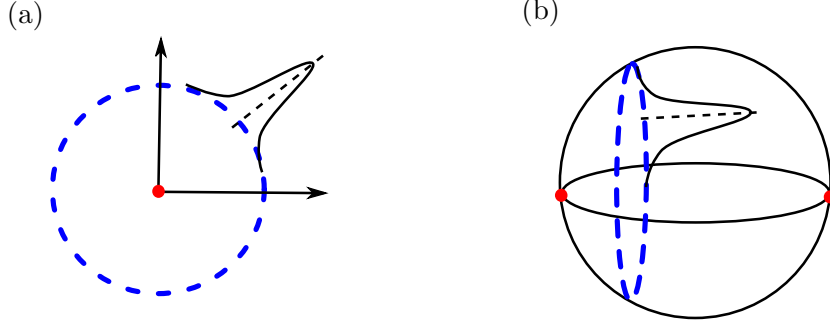


Figure 4. Step 2 of generative model of Equation (2) for the plane example (a) and the sphere example (b). The blue dotted curve illustrates the orbit of the shape drawn in Step 1. The black curve illustrates the probability distribution function on this orbit. This is a distribution in angle $\theta \in [0, 2\pi]$ for the plane example (polar coordinates) and in angle $\phi \in [0, 2\pi]$ for the sphere example (spherical coordinates).

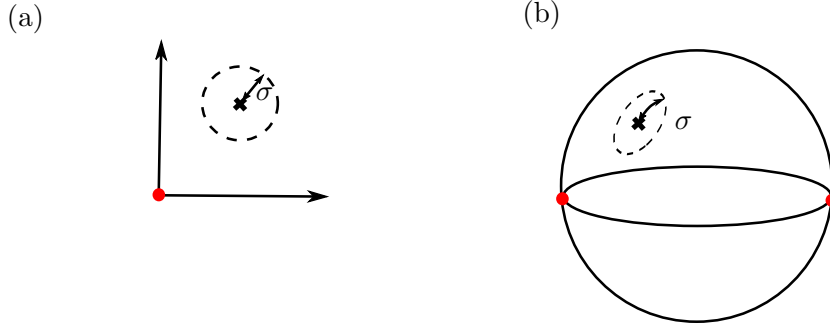


Figure 5. Step 3 of generative model of Equation (2) for the plane example (a) and the sphere example (b). The dotted curve represents the isolevel at σ of the Gaussian distribution function on the ambient space.

Estimating the template shape with the Fréchet mean in the shape space. We describe the procedure usually performed in the literature [25, 2, 4, 8, 7]. One initializes the estimate with $\hat{Y} = X_1$. Then, one iterates the following two steps until convergence:

$$(3) \quad \begin{cases} (i) & \hat{g}_i = \operatorname{argmin}_{g \in G} d_M(\hat{Y}, g \cdot X_i), \quad \forall i \in \{1, \dots, n\}, \\ (ii) & \hat{Y} = \operatorname{argmin}_{Y \in M} \sum_{i=1}^n d_M(Y, \hat{g}_i \cdot X_i)^2. \end{cases}$$

This procedure has a very intuitive interpretation. Step (i) is the projection of each object X_i in the shape space Q , as illustrated in Figure 6 (a)-(i) and (b)-(i) with the blue arrows. We assume that each minimizer \hat{g}_i exists and is attained. In practice, this is true for example when the Lie group is compact, like the Lie group of rotations. We take X_1, X_2, X_3 three objects in \mathbb{R}^2 in Figure 6 (a)-(i) and on S^2 in Figure 6 (b)-(i). One filters out the position/parameterization component, i.e. the coordinate on the orbit. One projects the objects X_1, X_2, X_3 in the shape space Q using the blue arrows.

Step (ii) is the computation of the mean of the registered data $\hat{g}_i \cdot X_i$, i.e. of the objects' shapes, as illustrated in Figure 6(a)-(i) and (b)-(i) where \hat{Y} is shown in orange. Again, we assume that the minimizer \hat{Y} exists and is attained. In practice, this will be the case as we will consider a low level of noise in Step 3 of the generative model. The registered data $\hat{g}_i \cdot X_i$ will be concentrated on a small neighborhood of diameter of order σ . As a consequence, their Fréchet mean in the Riemannian manifold Q^* is guaranteed to exist and be unique [17].

The procedure of Equations (3) (i)-(ii) decreases at each step the following cost, which is bounded below by zero:

$$(4) \quad \text{Cost}(g_1, \dots, g_n, Y) = \sum_{i=1}^n d_M^2(Y, g \cdot X_i).$$

Under the assumptions that both steps (i) and (ii) attained their minimizers, we are guaranteed convergence to a local minimum. We further assume that the procedure converges to the global minimum. The estimator computed with the procedure is then:

$$(5) \quad \hat{Y} = \underset{Y \in M}{\operatorname{argmin}} \sum_{i=1}^n \min_{g \in G} d_M^2(Y, g \cdot X_i).$$

The term $\min_{g \in G} d_M^2(Y, g \cdot X_i)$ in Equation 5 is the distance in the shape space between the shapes of Y and X_i . Thus, we recognize in Equation 5 the Fréchet mean on the shape space. The Fréchet mean is a definition of mean on manifolds [36]: it is the point that minimizes the squared distances to the data in the shape space. All in all, one projects the probability distribution function of the X_i 's from M to M/G and computes its "expectation", in a sense made precise later.

We implemented the generative model and the estimation procedure on the plane and the sphere in shiny applications available online: <https://nmiolane.shinyapps.io/shinyPlane> and <https://nmiolane.shinyapps.io/shinySphere>. We invite the reader to look at the web pages and play with the different parameters of the generative model. Figure 7 shows screen shots of the applications.

Probabilistic interpretation of the procedure in Equations (3): an approximation of a Maximum-Likelihood. Beside its intuitive interpretation, the procedure of template shape estimation of Equations (3) has a probabilistic interpretation. We have the generative model of the data X_i 's: it is described in Equation (2) and Steps (1)-(3) of the previous subsections. Thus, one may consider the Maximum Likelihood (ML) estimate of Y , which is one of its parameters:

$$\begin{aligned} \hat{Y}_{ML} &= \underset{Y \in Q}{\operatorname{argmax}} L(Y) = \underset{Y \in Q}{\operatorname{argmax}} \sum_{i=1}^n \log(P(X_i|Y)) \\ &= \underset{Y \in Q}{\operatorname{argmax}} \sum_{i=1}^n \log \left(\int_{g \in G} P(X_i|Y, g) \cdot P(g) dg \right). \end{aligned}$$

In the above, $P(X_i|Y)$ is the probability distribution of the data in M as a function of the parameter Y . $P(g)$ is the probability distribution on the poses/parameterizations in G as

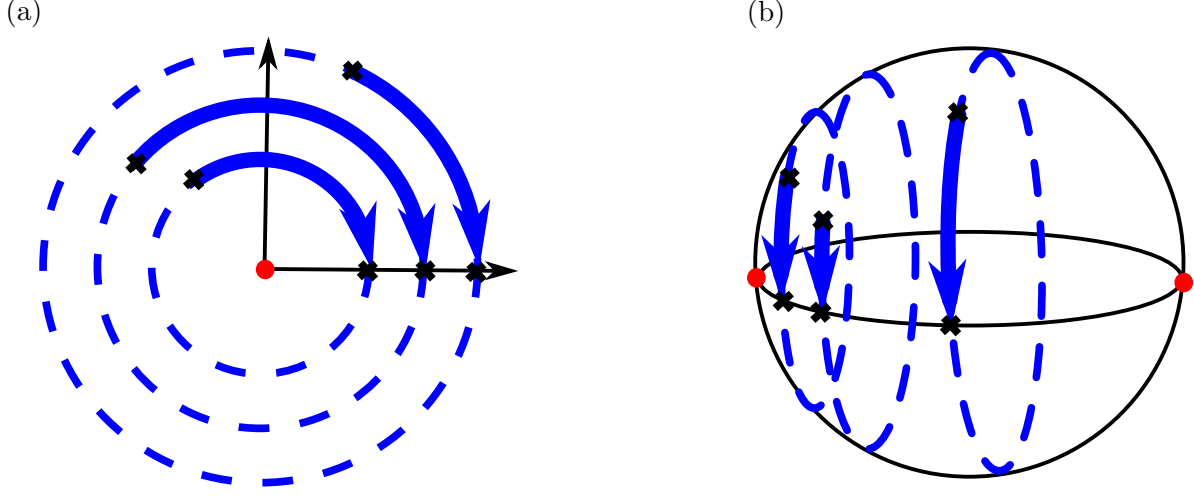


Figure 6. Steps (i) and (ii) of procedure of template shape estimation described in Equations (3) (i)-(ii) for the plane example (a) and the sphere example (b). The 3 black plus signs in \mathbb{R}^2 (a) or S^2 (b) represent the 3 data. The 3 dotted blue curves are their orbits. In Step (i), the X_i 's are registered, i.e. their projected in the shape space: 3 curved blue arrows represent their registration with the minimizers \hat{g}_i . The 3 black crosses in \mathbb{R}_+ (positive x-axis) (a) or $[0, \pi]$ (b) represent the registered data. In Step (ii), the template shape estimate \hat{Y} is computed as the Fréchet mean of the registered data and is shown in orange.

described in Step 2 of the generative model given in the previous subsection. Then, $P(X_i|Y, g)$ is the probability distribution of the noise as described in Step 3.

The g 's are hidden variables in the model. The Expectation-Maximization (EM) algorithm is therefore the natural implementation for computing the ML estimator [2]. But the EM algorithm is computationally expensive, above all for tridimensional images. Thus, one can usually rely on an approximation of the EM, which is described in [2] as the "modal approximation" and used in [3, 25, 8].

We can check that this approximation is the procedure described in Equations (3). Step (i) is an estimation of the hidden observations g_i and an approximation of the E-step of the EM algorithm. Step (ii) is the M-step of the EM algorithm: the maximization of the surrogate in the M-step amounts to the maximization of the variance of the projected data. This is exactly the minimization of the squared distances to the data of (ii). We refer to [2] for details.

Purpose of this paper reformulated with the geometrization. Our main result is to show that the procedure presented in Equations (3) (and illustrated on Figure 6) gives an asymptotically biased estimate \hat{Y} for the template shape Y of the generative model presented in Equation (2) (and illustrated in Figures 3, 4 and 5). Figures 7 (a)-(d) present what is meant by *asymptotic bias*: the estimate \hat{Y} , of the procedure, is in orange and the template shape Y , of the generative model, is in green. The estimator \hat{Y} (in orange) does converge when the number of data, i.e. the grey points in Figures 7(a)-(c), goes to infinity, *but \hat{Y} does not converge to the template shape Y it is designed to estimate*. For Figures 7 (a)-(d), this means that even for an infinite number of grey points, the orange estimate will be different from the green parameter. We

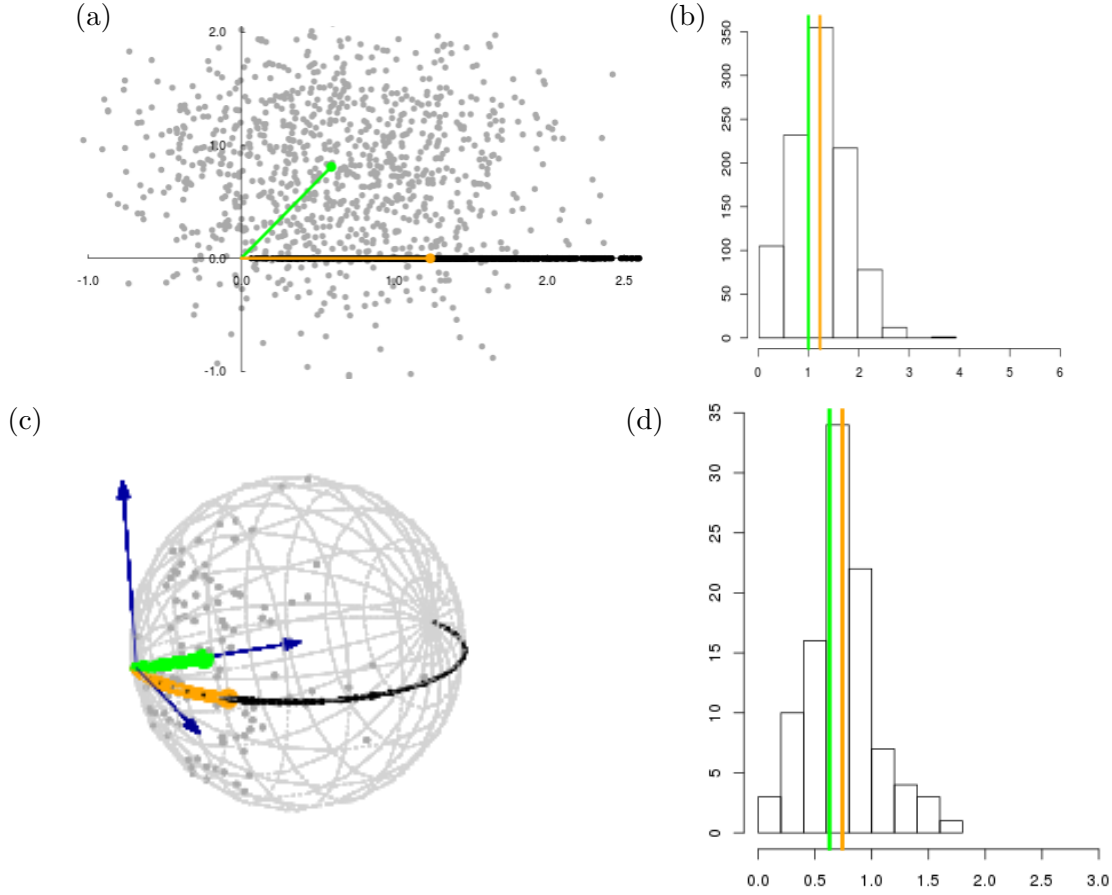


Figure 7. Screenshot of <https://nmiolane.shinyapps.io/shinyPlane> and <https://nmiolane.shinyapps.io/shinySphere>. Simulated data X_i 's (grey points), template shape Y (green), registered data $\hat{g}_i \cdot X_i$ (black points), template shape estimate \hat{Y} (orange). Induced distributions on the shapes, template shape Y (green), template shape estimate \hat{Y} (orange).

say that \hat{Y} has an asymptotic bias with respect to the parameter Y .

Where does this asymptotic bias come from and why doesn't \hat{Y} converge to Y ? In a nutshell, the bias comes from the external curvature of the template's orbit and we explain and summarize this in Figure 11 and its caption. The full geometric answer with its technical details is provided in the next section.

2. Quantification and correction of the asymptotic bias. This section explains, quantifies and corrects the asymptotic bias of the template shape estimate \hat{Y} with respect to the parameter Y . We start from the definition of the asymptotic bias of an estimator with respect to the parameter it is designed to estimate. More precisely we start from a generalization of this definition to Riemannian manifolds:

$$(6) \quad \text{Bias}(\hat{Y}, Y) = \mathbb{E} \left[\text{Log}_Y \hat{Y} \right].$$

This is the asymptotic bias of the estimator \hat{Y} with respect to the (manifold-valued) parameter Y , which generalizes the corresponding definition for linear spaces:

$$(7) \quad \text{Bias}(\hat{Y}, Y) = \mathbb{E} [\hat{Y} - Y]$$

In the Riemannian definition of the bias, $\text{Log}_Y \hat{Y}$ is the Riemannian logarithm of $\hat{Y} \in Q$ at $Y \in Q$, i.e. a vector of the tangent space of Q at the real parameter Y , denoted $T_Y Q$. The tangent vector $\text{Log}_Y \hat{Y}$ is illustrated on Figures 8 (a) and (b) for the plane and sphere examples. $\text{Log}_Y \hat{Y}$ represents how much one would have to shoot from Y to get the estimated parameter \hat{Y} . The norm of $\text{Log}_Y \hat{Y}$, computed using the metric of Q at Y , represents the dissimilarity between \hat{Y} and Y .

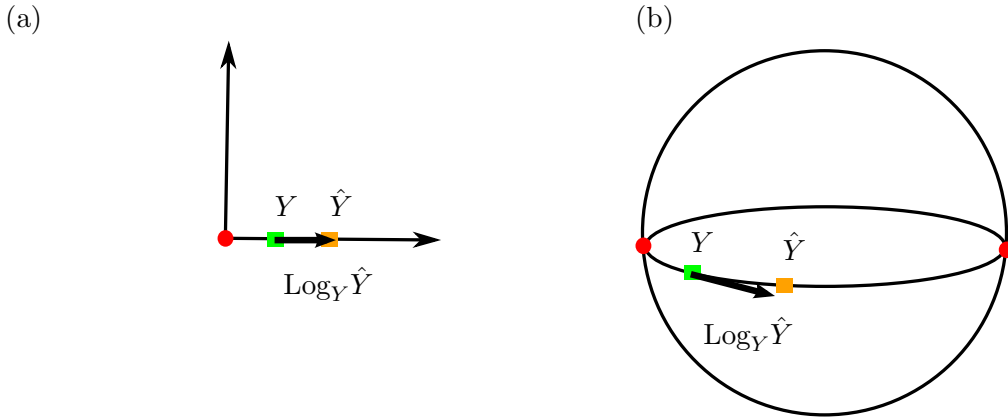


Figure 8. Illustration of the Riemannian definition of the asymptotic bias $\text{Log}_Y \hat{Y}$ for the plane example (a) and the sphere example (b). Log refers to the Riemannian logarithm [38] and $\text{Log}_Y \hat{Y}$ is thus a tangent vector of the quotient space Q at Y . $\text{Log}_Y \hat{Y}$ represents how much one would have to shoot from Y to get the estimated parameter \hat{Y} . The norm of $\text{Log}_Y \hat{Y}$, computed using the metric of Q at Y , represents the distance or the dissimilarity between \hat{Y} and Y , i.e. how far \hat{Y} is from estimating Y .

We could also consider the variance of the estimator \hat{Y} . The variance is defined as $\text{Var}_n(\hat{Y}) = \mathbb{E}[d_M(Y, E[Y])^2]$. In the limit of an infinite sample, we have: $\text{Var}_\infty(\hat{Y}) = 0$. This is why we focus on the asymptotic bias.

2.1. Asymptotic bias of the template's estimator on examples. We first compute the asymptotic bias for the examples of the plane and the sphere to give the intuition.

The probability distribution function of the X_i 's comes from the generative model. This is a probability distribution on \mathbb{R}^2 for the plane example, parameterized in polar coordinates (r, θ) like Figure 1. So we can compute the projected distribution function on the shapes, which are the radii r here. This is done simply by integrating out the distribution on θ , the position on the circles. This gives a probability distribution on \mathbb{R}_+ for the plane example. We write it $f : r \mapsto f(r)$. We remark that f does not depend on the probability distribution function on the θ_i 's of Step 2 of the generative model. We can also compute $f : \theta \mapsto f(\theta)$ in the sphere example: we integrate over ϕ the probability distribution function on (θ, ϕ) .

Figure 9 (a) shows f for the plane example, for a template $r = 1$. We plot it for two different noise levels $\sigma = 0.3$ and $\sigma = 3$. Note that here f is the Rice distribution. Figure 9 (b) shows f for the sphere example, for a template $\theta = 1$. We plot it for different noise levels $\sigma = 0.3$ and $\sigma = 3$. In both cases, the x-axis represents the shape space which is \mathbb{R}_+ for the plane example and $[0, \pi]$ for the sphere example. The green vertical bar represents the template shape, which is 1 in both cases. The red vertical bar is the expectation of f in each case. It is \hat{Y} , the estimate of Y . We see on these plots that f is not centered at the template shape: the green and red bars do not coincide. f is skewed away from 0 in the plane example and away from 0 and π in the sphere example. The skew increases with the noise level σ . The difference between the green and red bars is precisely the bias of \hat{Y} with respect to Y .

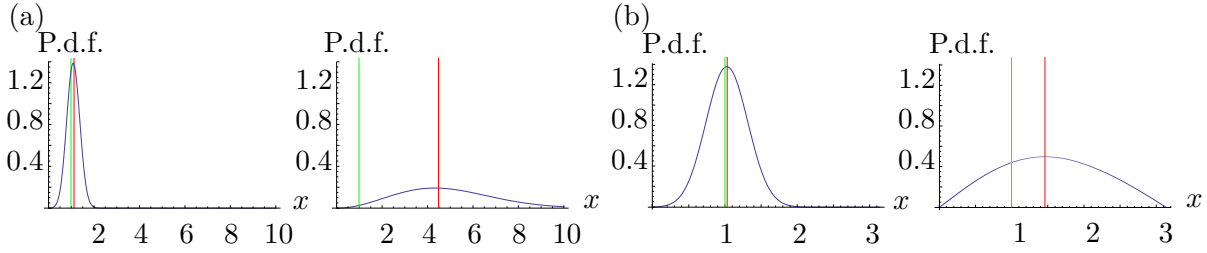


Figure 9. (a) Induced distributions on the distance r between two landmarks in \mathbb{R}^3 for real distance $y = 1$ (in green) and noise level $\sigma = 0.3$ and $\sigma = 3$. (b) Induced distributions on the angle x between the two landmarks on S^3 , for real angle $y = 1$ and noise levels $\sigma = 0.3$ and $\sigma = 3$. In both cases the mean shape estimate \hat{y} is shown in red.

Figure 10 shows the bias of \hat{Y} with respect to Y , as a function of σ , for the plane (left) and the sphere (right). Increasing the noise level σ takes the estimate \hat{Y} away from Y . The estimate is repulsed from 0 in the plane example: it goes to ∞ when $\sigma \rightarrow \infty$. It is repulsed from 0 and π in the sphere example: it goes to $\pi/2$ when $\sigma \rightarrow \pi$, as the probability distribution becomes uniform on the sphere in this limit. One can show numerically that the bias varies as σ^2 around $\sigma = 0$ in both cases. This is also observed on the shiny applications [39] at <https://nmiolane.shinyapps.io/shinyPlane> and <https://nmiolane.shinyapps.io/shinySphere>.

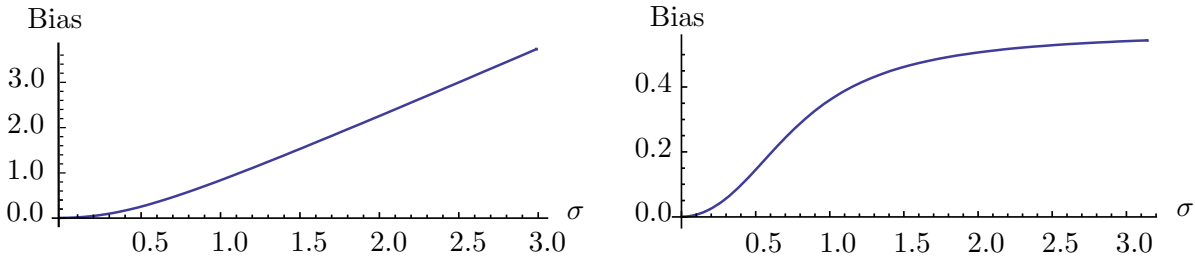


Figure 10. Asymptotic bias on the mean shape estimate \hat{Y} with respect to the noise level σ for $r = 1$ in the plane example (a) and $\theta = 1$ in the sphere example (b). The bias is quadratic near $\sigma = 0$. Increasing σ takes the estimate \hat{Y} away from 0 in shape space $Q = \mathbb{R}_+$ (a) and away from 0 and π in shape space $Q = [0, \pi]$ (b).

These examples already show the origin of the asymptotic bias of \hat{Y} , for low noise levels $\sigma \rightarrow 0$ or for high noise levels: $\sigma \rightarrow +\infty$ for the plane example and $\sigma \rightarrow \pi$ for the sphere

example. As long as there is noise, i.e. $\sigma \neq 0$, there is a *bias that comes from the curvature of the template's orbit*. Figure 11 shows the template's orbit in blue, in (a) for the plane and (b) for the sphere. In both cases the black circle represents the level set σ of the Gaussian noise. In the plane example (a), the probability of generating an observation X_i outside of the template's shape orbit is bigger than the probability of generating it inside: the grey area in the black circle is bigger than the white area in the white circle. There will be more registered data that are greater than the template. Their expected value will therefore be greater than the template and thus biased. In the sphere example (b), if the template's shape orbit is defined by a constant $\theta < \pi/2$, the probability of generating an observation X_i "outside" of it, i.e. with $\theta_i > \theta$, is bigger than the probability of generating it "inside". There will be more registered data that are greater than the template θ and again, their expected value will also be greater than the template. Conversely, if the template is $\theta > \pi/2$, the phenomenon is inversed: there will be more registered data that are smaller than the template. The average of these registered data will also be smaller than the template. Finally, if the template's shape orbit is the great circle defined by $\theta = \pi/2$, then the probability of generating an observation X_i on the left is the same as the probability of generating an observation X_i on the right. In this case, the registered data will be well-balanced around the template $\theta = \pi/2$ and their expected value will be $\pi/2$: there is no asymptotic bias in this particular case. We prove this in the general case in the next section.

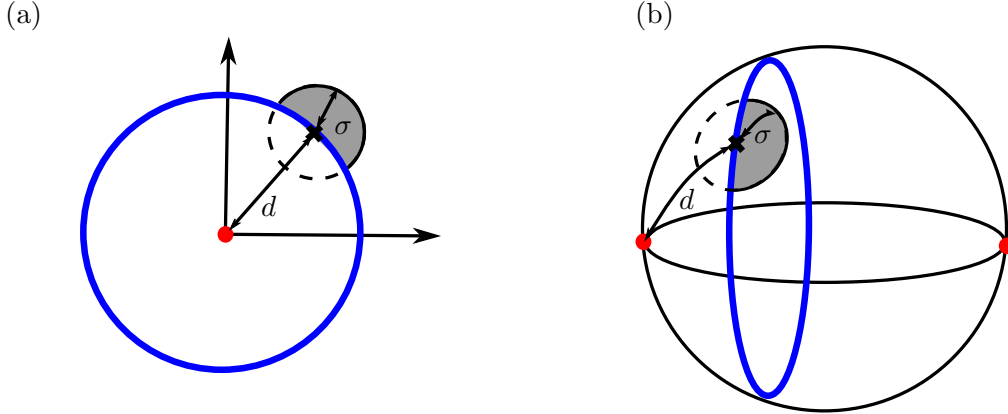


Figure 11. The external curvature of the template's orbit creates the asymptotic bias, in the plane example (a) and the sphere example (b). The blue curve represents the template's orbit. The ball of radius σ represents a level set of the Gaussian distribution of the noise in \mathbb{R}^2 (a) and S^2 (b). The grey-colored area represents the distribution of the noise that generates data outside the orbit of Y , in Step 3 of the generative model of Equation (2) and Figure 5. There is a higher probability that the data are generated "outside" the orbit. The template shape estimate is biased towards greater radii (a) or towards angles closer to $\pi/2$ (b).

2.2. Asymptotic bias of the template's estimator for the general case. We show the asymptotic bias of \hat{Y} in the general case and prove that it comes from the external curvature of the template's orbit. We show it for Y a principal shape and for a Gaussian noise of variance σ^2 , truncated at 3σ . Our results will need the following definitions of curvature.

The *second fundamental form* h of a submanifold O of M is defined on $T_X O \times T_X O$ by

$h(v, w) = (\nabla_v w)^\perp \in N_X O$, where $(\nabla_v w)^\perp$ denotes the orthogonal projection of covariant derivative $\nabla_v w$ onto the normal bundle. The *mean curvature vector* H of O is defined as: $H = \text{Tr}(h)$. Intuitively, h and H are measures of extrinsic curvature of O in M . For example an hypersphere of radius R in \mathbb{R}^m has mean curvature vector $\|H\| = \frac{m-1}{R}$.

Theorem 1. *The data X_i 's are generated in the finite-dimensional Riemannian manifold M following the model: $X_i = \text{Exp}(g_i \cdot Y, \epsilon_i)$, $i = 1 \dots n$, described in Equation (2) and Figures 3-5. In this model: (i) the action of the finite dimensional Lie group G on M , denoted \cdot , is isometric, (ii) the parameter Y is the template shape in the shape space Q , (iii) ϵ_i is the noise and follows a (generalization to manifolds of a) Gaussian of variance σ^2 , see Section 1.*

Then, the probability distribution function f on the shapes of the X_i 's, $i = 1 \dots n$, in the asymptotic regime on an infinite number of data $n \rightarrow +\infty$, has the following Taylor expansion around the noise level $\sigma = 0$:

$$f(Z) = \frac{1}{(\sqrt{2\pi}\sigma)^q} \exp\left(-\frac{d_M^2(Y, Z)}{2\sigma^2}\right) (F_0(Z) + \sigma^2 F_2(Z) + \mathcal{O}(\sigma^4) + \epsilon(\sigma))$$

where (i) Z denotes a point in the shape space Q , (ii) F_0 and F_2 are functions of Z involving the derivatives of the Riemannian tensor at Z and the derivatives of the graph G describing the orbit O_Z at Z , and (iii) ϵ is a function of σ that decreases exponentially for $\sigma \rightarrow 0$.

Proof. The sketch of the proof is given in Appendices, with the expressions of F_0 and F_2 . The detailed proof is in the supplementary materials. ■

The exponential in the expression of f belongs to a Gaussian distribution centered at Z and of isotropic variance $\sigma^2 \mathbb{I}$. However the whole distribution f differs from the Gaussian because of the Z -dependent term in the right parenthesis. This induces a skew of the distribution away from the singular shapes, as observed for the examples in Figure 9. This also means that the expectation of this distribution is not Z and that the variance is not the isotropic $\sigma^2 \mathbb{I}$.

Theorem 2. *The data X_i 's are generated with the model described in Equation (2) and Figures 3-5, where the template shape Y is a parameter and under the assumptions of Theorem 1. The template shape Y is estimated with \hat{Y} , which is computed by the usual procedure described in Equations (3).*

In the regime of an infinite number of data $n \rightarrow +\infty$, the asymptotic bias of the template's shape estimator \hat{Y} , with respect to the parameter Y , has the following Taylor expansion around the noise level $\sigma = 0$:

$$(8) \quad \text{Bias}(\hat{Y}, Y) = -\frac{\sigma^2}{2} H(Y) + \mathcal{O}(\sigma^4) + \epsilon(\sigma)$$

where (i) H is the mean curvature vector of the template shape's orbit which represents the external curvature of the orbit in M , and (ii) ϵ is a function of σ that decreases exponentially for $\sigma \rightarrow 0$.

Proof. The sketch of the proof is given in Appendices. The detailed proof is in the supplementary materials. ■

This generalizes the quadratic behavior observed in the examples on Figure 10. The asymptotic bias has a geometric origin: it comes from the external curvature of the template's orbits, see Figure 11.

We can vary two parameters in equation 8: Y and σ . The external curvature of orbits generally increases when Y is closer to a singularity of the shape space (see Section 1) [31]. The singular shape of the two landmarks in \mathbb{R}^2 arises when their distance is 0. In this case, the mean curvature vector has magnitude $|H(Y)| = \frac{1}{d}$: it is inversely proportional to d , the radius of the orbit. d is also the distance of Y to the singularity 0.

2.3. Limitations and extensions.

Beyond Y being a principal shape. Our results are valid when the template Y is a principal shape. This is a reasonable assumption as the set of principal shapes is dense in the shape space. What happens when Y approaches a singularity, i.e. when Y changes stratum in the stratified space Q ? Taking the limit $d \rightarrow 0$ in the coefficients of the Taylor expansion is not a legal operation. Therefore, we cannot conclude on the Taylor expansion of the Bias for $d \rightarrow 0$. Indeed, the Taylor expansion may even change order for $d \rightarrow 0$. We take $M = \mathbb{R}^m$ with the action of $SO(m)$ and the template $Y = (0, \dots, 0)$:

$$(9) \quad \text{Bias}(\hat{Y}, Y) = \sqrt{2} \frac{\Gamma(\frac{m+1}{2})}{\Gamma(\frac{m}{2})} \sigma.$$

The bias is linear in σ in this case.

Beyond $\sigma \ll 1$. The assumption $\sigma \ll 1$ represents our hope that the noise on the shape data is not too large with respect to the overall size of the mean shape. Nevertheless it would be very interesting to study the asymptotic bias for any σ , including large noises ($\sigma \rightarrow +\infty$). The distribution over the X_i 's in M will be spread on the whole manifold M . We cannot rely on local computations on M (at the scale of σ) anymore. We have to make global assumptions on the manifold M .

The plane example is the canonical example of a flat manifold. The sphere example is the canonical example of manifold with constant (positive) curvature. The bias as a function of σ is plotted in Figure 10. It leads us to the conjecture that the estimate converges towards a barycenter of shape space's singularities when the noise level increases. Singularities have a repulsive action on the estimation of each template's shape. Such repulsive force acts on each estimators. As a result, the estimators of the mean shape finds an equilibrium position: the barycenter.

Beyond one Dirac in Q : several templates. We have considered so far that there is a unique template shape Y : the generative model has a Dirac distribution at Y in the shape space. What happens for other distributions? We assume that there are K template shapes Y_1, \dots, Y_K . Observations are generated in M from each template shape Y_k with the generative model of Section 2. Our goal is to unveil the structure of the shape distribution, i.e. the K template shapes here, given the observations in M . The distributions on shapes projected on the shape space is a mixture of probability density functions of the form of equation 1. Its modes are related to the template shapes. The K-means algorithm is a very popular method for data clustering. We study what happens if one uses K-means algorithms on shapes generated with the generative model above.

The goal is to cluster the shape data in K distinct and significant groups. One performs

a coordinate descent algorithm on the following function:

$$(10) \quad J(c, \mu) = \sum_i d_Q(X_i, \mu_{c_i})^2.$$

In other words, the minimization of J is performed through successive minimizations on the assignment labels c 's and the cluster's centers μ 's. Given the c , minimizing J with respect to the μ 's is exactly the simultaneous computation of K Fréchet means in the shape space. Meaningful well separated clusters (high inter-clusters dissimilarity) are chosen so that members are close to each other (high intra-cluster similarity). In other words, the quality of the clustering is evaluated by the following criterion:

$$(11) \quad D = \min_{\text{clusters } i, j} \frac{d_Q(c_i, c_j)}{\max_i \text{diam}(c_i)},$$

which is the dissimilarity between clusters quotiented by the diameter of the clusters. In the absence of singularity in the shape space, the projected distribution looks like Figure 12 (a) and $D \propto \frac{1}{\sigma}$. The criterion is worse in the presence of singularities.

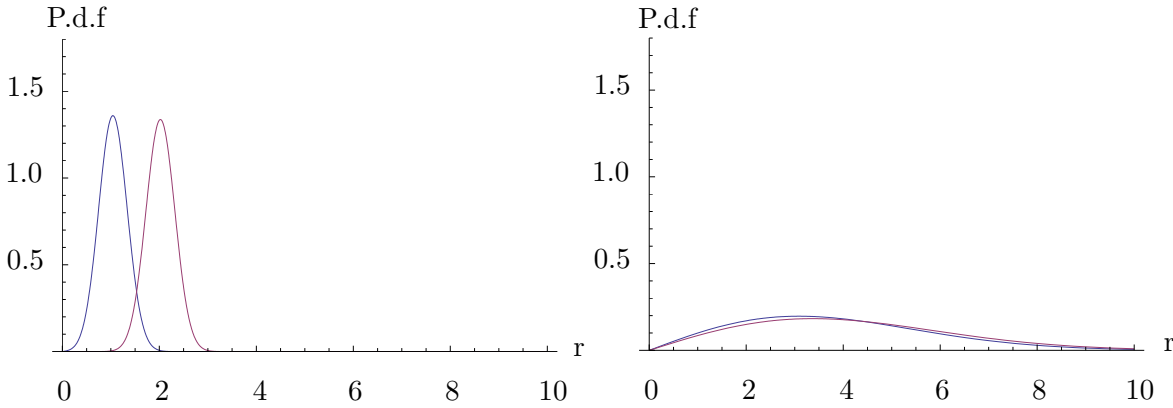


Figure 12. Two clusters of template shapes for the plane example: $r_1 = 1$ (blue) and $r_2 = 2$ (dark red). Noise levels: $\sigma = 0.3$ (left) and $\sigma = 3$ (right). The 2 clusters are hardly distinguishable when the noise increases.

Figure 12 illustrates this behavior for the plane example. We consider any two clusters i, j and call \hat{Y}_i, \hat{Y}_j the estimated centroids. The criterion D writes:

$$D \equiv \frac{\hat{y}_i - \hat{y}_j}{\sigma} \underset{\sigma \rightarrow +\infty}{\sim} \frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{2m}\Gamma\left(\frac{m}{2}\right)} \frac{y_i^2 - y_j^2}{\sigma^2} = O\left(\frac{1}{\sigma^2}\right).$$

Even in the best case with correct assignments to the clusters i and j , the K-means algorithm looses an order of validation when computed on shapes.

Beyond the finite dimensional case. Our results are valid when M is a finite dimensional manifold and G a finite dimensional Lie group. Some interesting examples belong to the framework of infinite dimensional manifold with infinite dimensional Lie groups. This is the case for the LDDMM framework on images [22]. It would be important to extend these results to the infinite dimensional case.

We take $M = \mathbb{R}^m$ with the action of $SO(m)$. We have an analytic expression of f in this case [33]. Figure 13 shows the influence of the dimension m for the probability distribution functions on the shape space and for the Bias. The bias increases with m . This leads us to think that it appears in infinite dimensions as well.

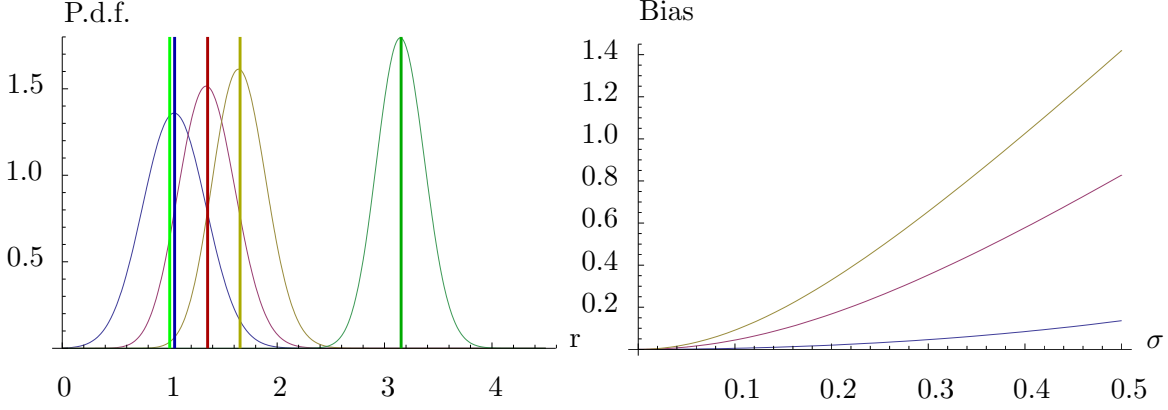


Figure 13. Probability distributions functions (noise $\sigma = 0.3$) and bias for \mathbb{R}^m for $m = 2$, $m = 10$, $m = 20$ and $m = 100$. Template shape is $r = 1$.

3. Correction of the systematic bias. We propose two procedures to correct the asymptotic bias on the template's estimate. They rely on the bootstrap principle, more precisely a parametric bootstrap, which is a general Monte Carlo based resampling method that enables us to estimate the sampling distributions of estimators [14]. We assume that we know the variance $\hat{\sigma}^2$ from the experimental setting.

3.1. Iterative Bootstrap. The first procedure is called an Iterative Bootstrap. Algorithm 3.1 gives the details. Figure 14 illustrates it on the plane example.

Algorithm 3.1 starts with the usual template's estimate $\hat{Y}_0 = \hat{Y}$, see Figure 14 (a). At each iteration, we correct \hat{Y} with a better approximation of the bias. First, we generate bootstrap data by using \hat{Y} as the template shape of the generative model. We perform the template's estimation procedure with the Fréchet mean in the shape space. This gives an estimate \hat{Y}_0^* of \hat{Y}_0 . The bias of \hat{Y}_0^* with respect to \hat{Y}_0 is $\text{Bias}(\hat{Y}_0^*, \hat{Y}_0)$. It gives an approximation of the bias $\text{Bias}(\hat{\hat{Y}}, \hat{Y})$, see Figure 14 (b). We correct \hat{Y} by this approximation of the bias. This gives a new estimate \hat{Y}_1 , see Figure 14 (c). We recall that the bias $\text{Bias}(\hat{\hat{Y}}, \hat{Y})$ depends on Y , see Theorem 2. \hat{Y}_1 is closer to the template Y than \hat{Y}_0 . Thus, the next iteration gives a better approximation $\text{Bias}(\hat{Y}_1^*, \hat{Y}_1)$ of $\text{Bias}(\hat{\hat{Y}}, \hat{Y})$. We correct the initial \hat{Y} with this better approximation of the bias, etc. The procedure is written formally for a general manifold M

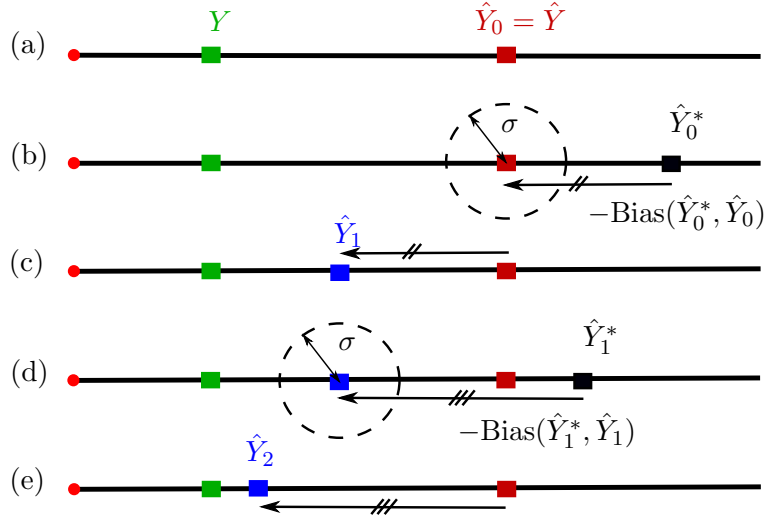


Figure 14. Algorithm 3.1 Iterative bootstrap procedure on the plane example for $n \rightarrow +\infty$. (a) Initialization, (b) Generate bootstrap sample from \hat{Y}_0 and compute the corresponding estimate \hat{Y}_0^* , compute the bias $\hat{Y}_0 - \hat{Y}_0^*$, (c) Correct \hat{Y}_0 with the bias to get \hat{Y}_1 , (d) Generate bootstrap sample from \hat{Y}_1 and iterate as in (b), (e) Get \hat{Y}_2 etc.

in Algorithm 3.1.

Algorithm 1 Corrected template shape estimation with **Iterative Bootstrap**

Input: Objects $\{X_i\}_{i=1}^n$, noise variance σ^2

Initialization:

$$\hat{Y}_0 = \text{Frechet}(\{[X_i]\}_{i=1}^n)$$

$$k \leftarrow 0$$

Repeat:

Generate bootstrap sample $\{X_i^{(k)*}\}_{i=1}^n$ from $\mathcal{N}_M(Y_k, \sigma^2)$

$$\widehat{Y}_k = \text{Fréchet}(\{[X^{(k)*}]_i\}_{i=1}^n)$$

$$\text{Bias}_k = \text{Log}_{Y_k} \widehat{Y}_k$$

$$\hat{Y}_k = \text{Exp}_{\hat{Y}_0} \left(-\Pi_{\hat{Y}_k}^{\hat{Y}_0} (\text{Bias}_k) \right)$$

$$k \leftarrow k + 1$$

until convergence: $\|\text{Log}_{\hat{Y}_{k+1}} \hat{Y}_k\| < \epsilon$

Output: \hat{Y}_k

In Algorithm 3.1, Π_A^B denotes the parallel transport from $T_A M$ to $T_B M$. For linear spaces like \mathbb{R}^2 in the plane example, $\text{Log}_{P_1} P_2 = \overrightarrow{P_1 P_2}$, $\text{Exp}_{P_1}(u) = P_1 + u$, and the parallel transport is the identity $\Pi_{P_1}^{P_2}(u) = u$. For other manifolds like S^2 in our sphere example, the parallel transport $\Pi_A^B(u)$ can theoretically be computed by solving the parallel transport equation at any point on the chosen curve linking A to B : $D_{t_{AB}} v = 0$ in v , where D is the covariant derivative in the direction t_{AB} , the tangent vector of the curve at the chosen point

[38]. In practice, the Schild's ladder [15] or the Pole ladder [32] can be used to compute an approximation of the parallel transport.

Algorithm 3.1 is a fixed-point iteration $Y^{(k+1)} = F(Y^{(k)})$ where:

$$(12) \quad F(X) = \text{Exp}_{\hat{Y}}(-\Pi_X^{\hat{Y}}(\text{Bias})) \quad \text{where:} \quad \text{Bias} = \text{Log}_X \hat{X}.$$

In a linear setting we have simply $F(X) = \hat{Y} - \overrightarrow{X\hat{X}}$. One can show that F is a contraction and that Y , the template shape, is the unique fixed point of F (using the local bijectivity of the Riemannian exponential and the injectivity of the estimation procedure). Thus the procedure converges to Y in the case of an infinite number of observations $n \rightarrow +\infty$. Figure 15 illustrates the convergence for the plane example, with a Gaussian noise of standard deviation $\sigma = 1$. The template shape $Y = 1.2$ was initially estimated at $\hat{Y} = 4.91$. Algorithm 3.1 corrects the bias.

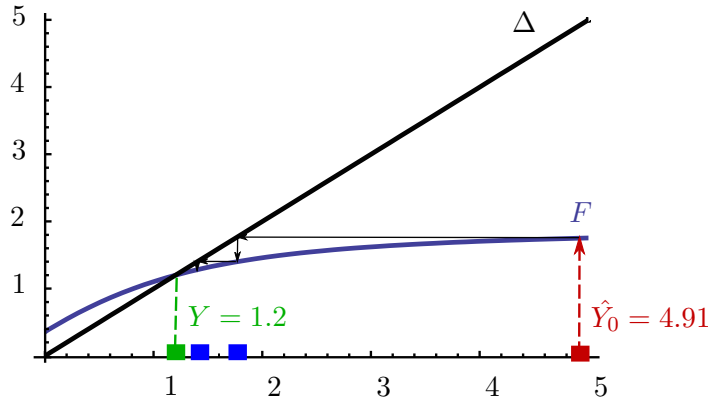


Figure 15. F of the fixed-point procedure and first 2 iterations for $\sigma = 1$, $m = 3$. Δ is the first diagonal. The initial estimate is biased $\hat{Y}_0 = 4.91$. The Iterative Bootstrap converges towards the template shape $Y = 1.2$.

Figures 16 and 17 show the iterations of Iterative Bootstrap for the plane and the sphere example.

3.2. Nested Bootstrap. The second procedure is called the Nested Bootstrap. Algorithm 3.2 details it. Figure 18 illustrates it on the plane example.

Algorithm 3.2 starts like Algorithm 3.1 with $\hat{Y}_0 = \hat{Y}$, see Figure 18 (a). It also performs a parametric bootstrap with $\hat{Y}_0^{(0)}$ as the template, computes the bootstrap replication \hat{Y}_0^* and the approximation $\text{Bias}(\hat{Y}_0^*, \hat{Y}_0)$ of $\text{Bias}(\hat{Y}, Y)$, see Figure 14 (b). Now Algorithm 3.2 differs from Algorithm 3.1. We want to know how biased is $\text{Bias}(\hat{Y}_0^*, \hat{Y}_0)$ as an estimate of $\text{Bias}(\hat{Y}, Y)$? This is a valid question as the bias depends on the template Y , see Theorem 2. We want to estimate this dependence. We perform a bootstrap, nested in the first one, with $\hat{Y}_0^{(0)*}$ as the template. We compute the estimate \hat{Y}_0^{**} and the approximation $\text{Bias}(\hat{Y}_0^{**}, \hat{Y}_0^*)$ of $\text{Bias}(\hat{Y}_0^*, \hat{Y}_0)$, see Figure 14 (c). We observe how far $\text{Bias}(\hat{Y}_0^{**}, \hat{Y}_0^*)$ is from $\text{Bias}(\hat{Y}_0^*, \hat{Y}_0)$. This gives the blue arrow, which is the bias of $\text{Bias}(\hat{Y}_0^{**}, \hat{Y}_0^*)$ as an estimate of $\text{Bias}(\hat{Y}_0^*, \hat{Y}_0)$, see Figure 14 (d). The blue arrow is an approximation of how far $\text{Bias}(\hat{Y}_0^*, \hat{Y}_0)$ is from $\text{Bias}(\hat{Y}, Y)$. We correct our estimation of the bias (in red) by the blue arrow. We correct \hat{Y} by the bias-corrected estimate of its bias, see Figure 14 (e).

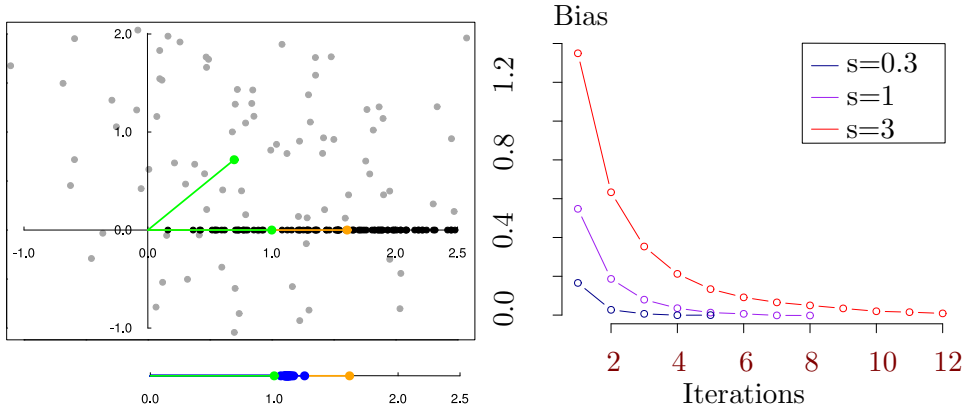


Figure 16. Left: Implementation of the plane example: the green point is the template shape Y , the grey points are the data X_i 's generated with the model (2), the black points are the registered data $\hat{g}_i \cdot X_i$'s, the orange point is the template shape estimate \hat{Y} . The quotient space \mathbb{R}_+ is copied below, and the blue points show the iterations of the iterative bootstrap of Algorithm 3.1 that corrects the bias of \hat{Y} as an estimate of Y : the blue points go from the orange point \hat{Y} to the green point Y . Right: Convergence of the iterative bootstrap of Algorithm 3.1, for the plane example. The colors red, purple, blue represent different noises σ . The bias of \hat{Y} as an estimator of Y is shown on the ordinate axis: it converges to 0 in a few iterations.

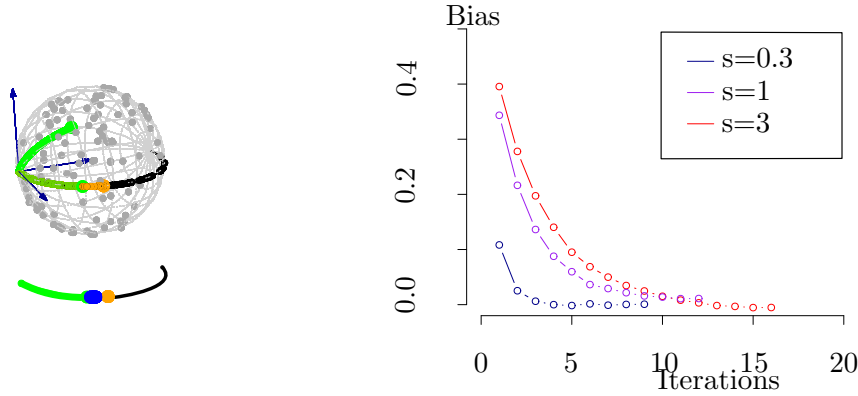


Figure 17. Left: Implementation of the sphere example: the green point is the template shape Y , the grey points are the data X_i 's generated with the model (2), the black points are the registered data $\hat{g}_i \cdot X_i$'s, the orange point is the template shape estimate \hat{Y} . The quotient space $[0, \pi]$ is copied below, and the blue points show the iterations of the iterative bootstrap of Algorithm 3.1 that corrects the bias of \hat{Y} as an estimate of Y : the blue points go from the orange point \hat{Y} to the green point Y . Right: Convergence of the iterative bootstrap of Algorithm 3.1, for the sphere example. The colors red, purple, blue represent different noises σ . The bias of \hat{Y} as an estimator of Y is shown on the ordinate axis: it converges to 0 in a few iterations.

3.3. Comparison. One may use the Iterative Bootstrap or the Nested Bootstrap depending on the experimental setting. We illustrate them both on the plane example in Figure 19. Figure 19 (a) shows the performance of both algorithms for a signal-over-noise ratio (SNR) of 1: the template shape in green is a $r = 1$ and the standard deviation of the noise is $\sigma = 1$, so that $\text{SNR} = \frac{r}{\sigma} = 1$. Figure 19 (b) shows both algorithms for $\text{SNR} = \frac{r}{\sigma} = \frac{1}{3} = 0.33$. In all four experiments: the template shape is the green dot at $r = 1$, the template shape estimate is in orange, and the successive steps of the bootstrap algorithms are the blue dots: we have

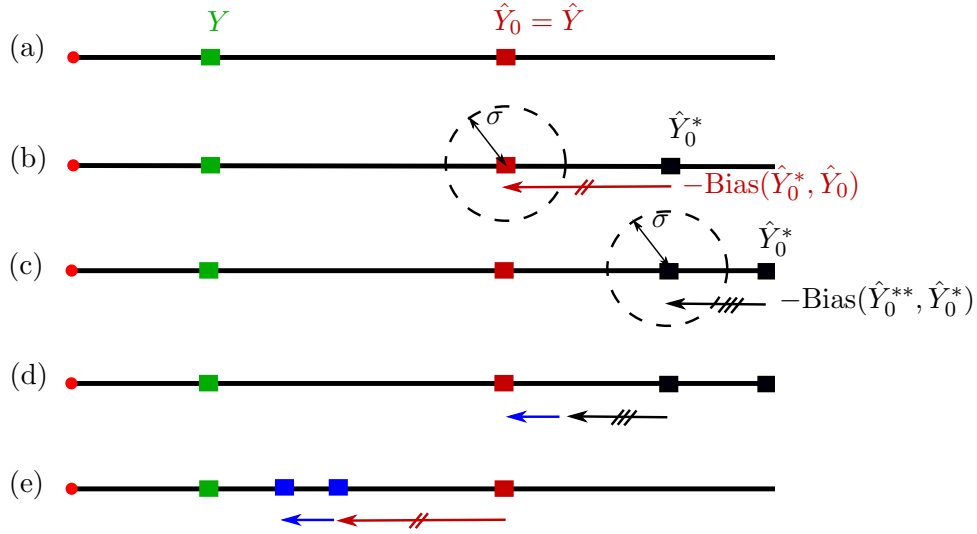


Figure 18. Algorithm 3.2 Nested Bootstrap on the plane example for $n \rightarrow +\infty$. (a) Initialization, (b) Generate bootstrap sample from \hat{Y}_0 ; compute the estimate \hat{Y}_0^* , compute the bias $\hat{Y}_{(0)} - \hat{Y}_0^*$, (c) Generate bootstrap sample from \hat{Y}_0^* ; compute the estimate \hat{Y}_0^{**} , compute the bias $\hat{Y}_0^* - \hat{Y}_0^{**}$, (d) compute the blue arrow, i.e. the bias of $\text{Bias}(\hat{Y}_0^{**}, \hat{Y}_0^*)$ as an estimate of $\text{Bias}(\hat{Y}_0^*, \hat{Y}_0)$, (e) Correct \hat{Y} with the bias-corrected bias.

Algorithm 2 Corrected template shape estimation with **Nested Bootstrap**

Input: Objects $\{X_i\}_{i=1}^n$, noise variance σ^2

Initialization:

$$\hat{Y}_0 = \text{Frechet}(\{[X_i]\}_{i=1}^n)$$

Bootstrap:

Generate bootstrap sample $\{X_i^*\}_{i=1}^n$ from $\mathcal{N}_M(\hat{Y}_0, \sigma^2)$

$$\hat{Y}_0^* = \text{Fréchet}(\{[X^*]_i\}_{i=1}^n)$$

$$\text{Bias} = \text{Log}_{\hat{Y}_0} \hat{Y}_0^*$$

Nested Bootstrap:

For each i :

- Generate bootstrap sample $\{X_{i,k}^{**}\}_{k=1}^n$ from $\mathcal{N}_M(\hat{Y}_0^*, \sigma^2)$
- $\hat{Y}_{0,i}^{**} = \text{Fréchet}(\{[X_{i,k}^{**}]\}_{k=1}^n)$

$$\text{Bias}(\text{Bias}) = \text{Log}_{\hat{Y}_0} \hat{Y}_0^* - \Pi_{\hat{Y}_0^*}^{\hat{Y}_0} \text{Log}_{\hat{Y}_0^*} \hat{Y}_{0,i}^{**}$$

$$\hat{Y}_1 = \text{Exp}_{\hat{Y}_0}(-\text{Bias} - \text{Bias}(\text{Bias}))$$

Output: \hat{Y}_1

several blue dots for the Iterative Bootstrap, and two blue dots for the Nested Bootstrap.

The advantages of the Iterative Bootstrap are the following. It corrects the bias of \hat{Y} perfectly in the case of a very large number of observations n , as we can see in Figures 19 (a) on top and (b) on top: the blue dots converge to the green dot for the two different SNRs. Thus, the Iterative Bootstrap can be used to experimentally compute the mean curvature vector H of each orbit of a group action. One probes the orbit's curvature by "feeling it"

with a Riemannian Gaussian on M and projecting on the shape space. The drawbacks of the Iterative Bootstrap are the following. It works only with very large n . It is not robust as it uses the generative model several times. If the generative model is far from being true, then the iterative bootstrap fails.

The advantages of the Nested Bootstrap are the following. It is a standard statistical procedure that is more robust with respect to variations of the generative model. Even if generative model is different from the one that we assume, the Nested Bootstrap performs well. Moreover, it does not need as much data as the Iterative Bootstrap. Its drawback is that it does not correct perfectly the bias, especially when the noise is large. This can be seen in Figures 19 (a) on bottom and (b) on bottom. While the Nested Bootstrap gets close to the green dot on Figure 19 (a) on bottom for the $SNR = 1$, it stays significantly far from the green dot on Figure 19 (b) bottom for the $SNR = 0.33$.

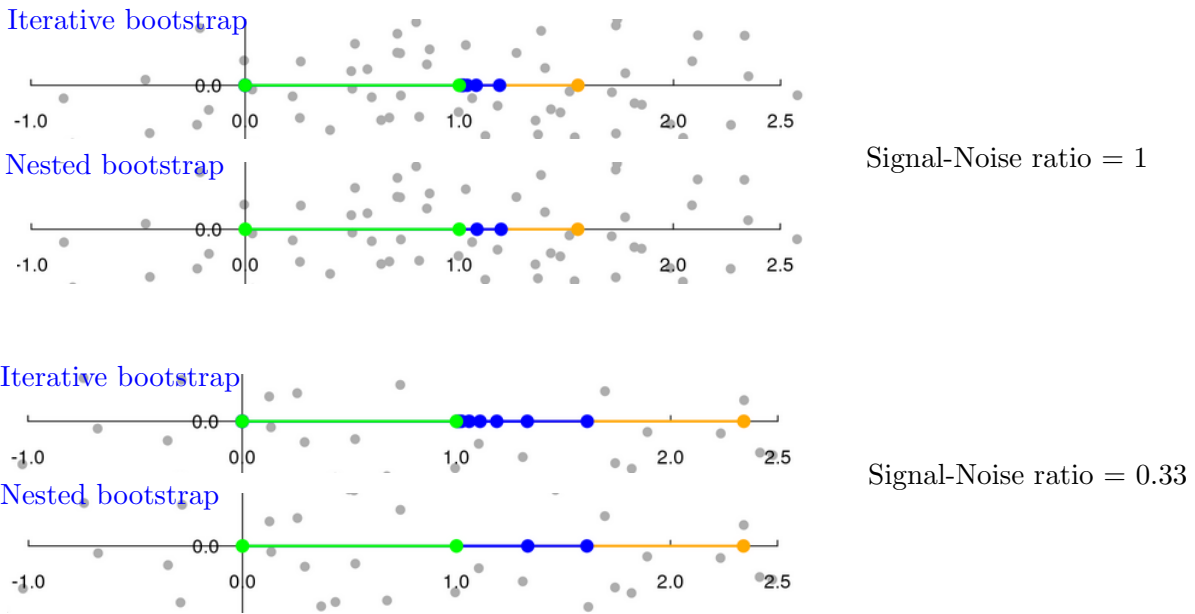


Figure 19. Comparison of the Iterative bootstrap and the Nested bootstrap on simulation with two different Signal-Noise ratio, which is $SNR = \frac{Y}{\sigma} = \frac{r}{\sigma}$, the ratio of the template Y , which is the radius r in the plane example, on the noise level σ . (a) shows $SNR = 1$ and (b) shows $SNR = 0.33$. In all four experiments: the template shape is the green dot at $r = 1$, the template shape estimate is in orange, and the successive steps of the bootstrap algorithms are the blue dots: we have several blue dots for the Iterative Bootstrap, and two blue dots for the Nested Bootstrap.

These simulations give a rule of thumb, i.e. some intuition, for when the bias needs to be corrected. They confirm what could already be observed in Figure 10. In Figure 10, the template is fixed at $r = 1$ or $\theta = 1$. A variation in the noise level σ corresponds to a variation in the SNR. In particular, we read the threshold $SNR = 1$ when $\sigma = 1$, i.e. when the noise σ is comparable to the distance of the template Y to the singularity. In both cases for $SNR > 1$, the template estimate is significantly different from the template as the bias is of the order of magnitude of the template itself.

4. Applications to simulated and real data.

4.1. Simulated triangles. We perform a simulation using the iterative bootstrap on triangles. We randomly generate $n = 10^5$ triangles in \mathbb{R}^2 through the generative model described in Equation (2) of Section 1. This is illustrated on Figure 20. We consider the isometric action of the Lie group $SO(2)$ of 2D rotations on $(\mathbb{R}^2)^3$, the space of 3 landmarks in 2D. For Step 1 of the generative model of Section 1, the template triangle is chosen arbitrarily and then fixed during the simulations. The template triangle is represented in green in Figure 20. For Step 2, we consider a Dirac distribution at the identity in the Lie group $SO(2)$. In other words, we do not rotate the triangles. At the end of this step, each of the 10^5 triangles is exactly the green triangle of Figure 20. This simpler model does not decrease the impact of the simulation: the noise of Step 3 is independent of the position of the triangle on their orbit, and Step (i) of the procedure is to quotient out the position of the orbit. For Step 3, we add bivariate Gaussian noise on each landmark, i.e. on each of the three points defining the green triangle. This gives a data set of 10^5 triangles. Some of them are represented in grey on Figure 20.

We then apply the procedure described in Equations (3) (i)-(ii) to estimate the (green) template triangle. In Step (i), we register the (grey) triangle data. This gives the registered data, illustrated in black in Figure 20. We then compute the Fréchet mean of the black triangles by computing the Euclidean mean of each of their 3 landmarks. This gives the estimate of the template triangle, in orange on Figure 20.

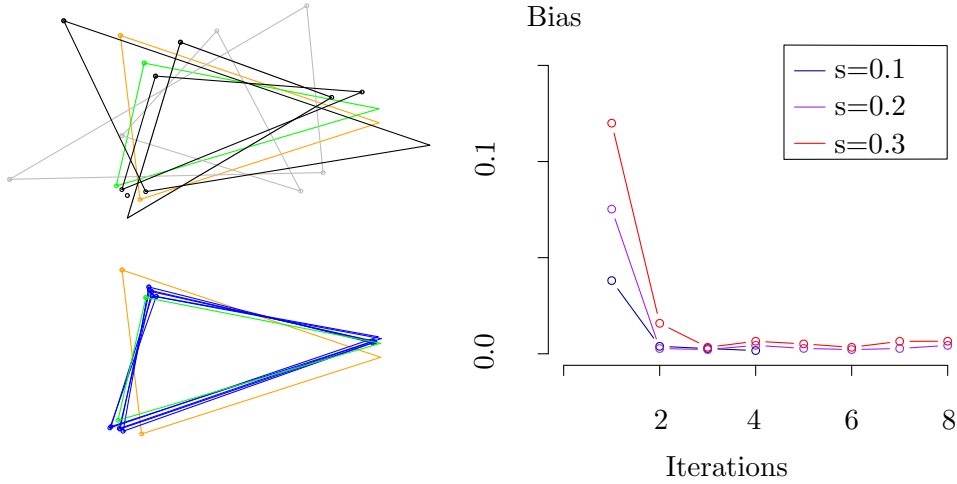


Figure 20. Left: Implementation for simulated triangles: the green triangle is the template shape Y , the grey triangles are (some of) the data X_i 's generated with the model (2), the black triangles are (some of) the registered data $\hat{g}_i \cdot X_i$'s, the orange triangle is the template shape estimate \hat{Y} . The blue triangles show the iterations of the iterative bootstrap of Algorithm 3.1 that corrects the bias of \hat{Y} as an estimate of Y : the blue triangles go from the orange triangle \hat{Y} to the green triangle Y . Right: Convergence of the iterative bootstrap of Algorithm 3.1. The colors red, purple, blue represent different noises σ . The bias of \hat{Y} as an estimator of Y is shown on the ordinate axis: it converges to 0 in a few iterations.

The template estimate in orange is different from the template in green, even with a very high number of observations: $n = 10^5$. We apply the iterative bootstrap to correct this bias. The number of iterations required for the convergence of Algorithm 1 with respect to the noise

level are shown in Figure 20. We observe the convergence in the three experiments for less than 10 iterations.

4.2. Real triangles: shape of the Optic Nerve Head. Now we go to real triangle data. We have 24 images of Rhesus monkeys' eyes, acquired with a Heidelberg Retina Tomograph [34]. For each monkey, an experimental glaucoma was introduced in one eye, while the second eye was kept as control. One seeks a significant difference between the glaucoma and the control eyes. On each image, three anatomical landmarks were recorded: S for the superior aspect of the retina, N for the nose side of the retina, and T for the side of the retina closest to the temporal bone of the skull. The data are matrices $\{X_i\}_{i=1}^n$ where the landmark coordinates form the rows. For the ONH example, M is the space of 3 landmarks in 3D, $M = (\mathbb{R}^3)^3$ and the rotations act isometrically on each object X_i .

Analysis This simple example illustrates the estimation of the template shape. We use the following procedure to compute the mean shape for each group. We initialize \hat{Y} with X_1 and repeat the following two steps until convergence:

- (1) $\forall i \in \{1, \dots, n\}, \quad \hat{R}_i = \underset{R \in SO(3)}{\operatorname{argmin}} \|\hat{Y} - X_i \cdot R\|^2, \quad (\text{register to the current mean shape}),$
- (2) $\hat{Y} = \frac{1}{n} \sum_{i=1}^n X_i \cdot \hat{R}_i \quad (\text{update the mean shape estimate}).$

Figure 21 shows the mean shapes \hat{Y}^{control} of the control group (left) and $\hat{Y}^{\text{glaucoma}}$ of the glaucoma group (right) in orange, while the initial data are in grey. The difference between the two groups is quantified by the distance between their means: $\|\hat{Y}^{\text{control}} - \hat{Y}^{\text{glaucoma}}\| = 21.84\mu\text{m}$. We want to determine if this analysis presents a bias that significantly changes the estimated shape difference between the groups.

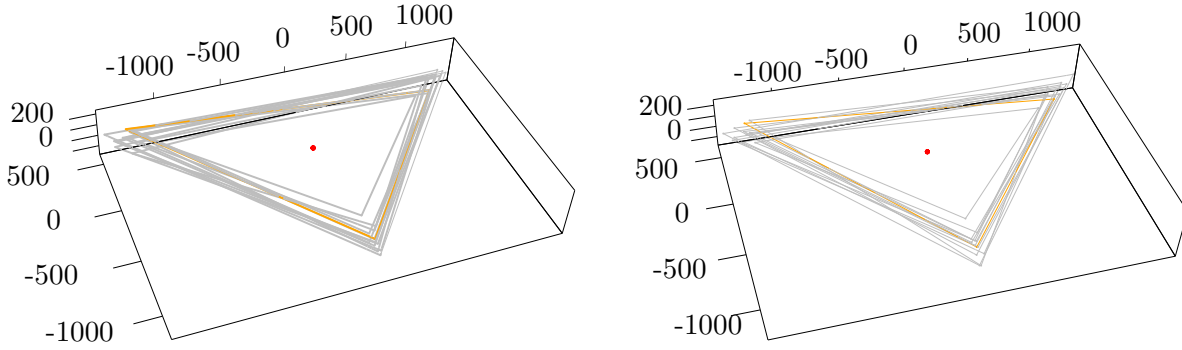


Figure 21. Triangles data in grey for the control group (left) and the glaucoma group (right). In orange, the estimated template shapes. Distances are measured in μm .

We use the nested bootstrap to compute an approximation of the asymptotic bias on each mean shape, for a range of noise's standard deviation in $\{100\mu\text{m}, 200\mu\text{m}, 300\mu\text{m}, 400\mu\text{m}\}$. The asymptotic bias on the template shape of the glaucoma group is $\{0.1\mu\text{m}, 0.11\mu\text{m}, 0.12\mu\text{m}, 0.13\mu\text{m}\}$ and of the control group is $\{0.27\mu\text{m}, 0.42\mu\text{m}, 0.55\mu\text{m}, 0.67\mu\text{m}\}$. The corrected template shape differences are $\{22.01\mu\text{m}, 22.08\mu\text{m}, 22.14\mu\text{m}, 22.18\mu\text{m}\}$. In particular, for $\sigma =$

400 μm , we observe that the bias in the template shape are respectively 0.67 μm for the healthy group and 0.13 μm for the glaucoma group. This follows the rule-of-thumb: the bias is more important for the healthy group, for which the overall size is smaller than the glaucoma group, for a same noise level. The bias of the template shape estimate accounts for less than 1 μm in this case, which is less than 0.1% of the shapes' sizes. This computation guarantees that this study has not been significantly affected by the bias.

4.3. Protein shapes in Molecular Biology. We estimate the impact of the bias on statistics on protein shapes. This subsection aims to suggest the potential importance of the results of this paper for Molecular Biology.

A standard hypothesis in Biology is that structure (i.e. shape) and function of proteins are related. Fundamental research questions about protein shapes include structure prediction - given the protein amino-acid sequence, one tries to predict its structure - and design - given the shape, one tries to predict the sequence needed.

One relies on experimentally determined 3D structures gathered in the Protein Data Base (PDB) [6]. They contain errors on the protein's atoms coordinates. Average errors range from 0.01 Å to 1.76 Å, which is of the magnitude of the length of some covalent bonds. These values are averaged over the whole protein and in general, the main-chain atoms are better defined than the side-chain atoms or the atoms at the periphery. This is illustrated on Figure 22 where we have plot the B-factor (related to coordinates errors [40]) as a colored map on the atoms for proteins of PDB-codes 1H7W and 4HBB.

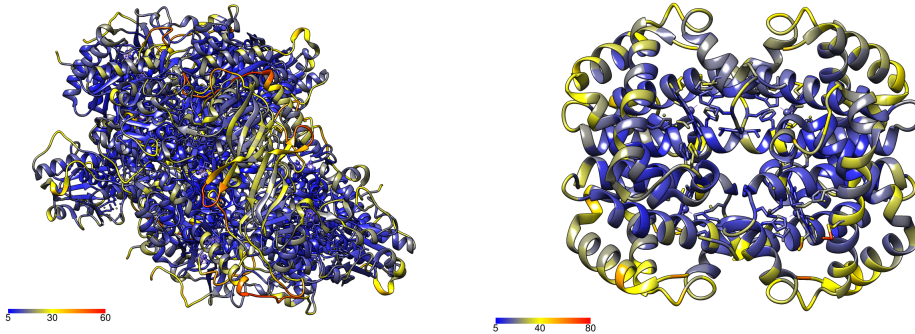


Figure 22. Errors on atoms coordinates represented by the B-factor, for proteins 1H7W (left) and 4HBB (right). Atoms at the periphery of the proteins tend to have more errors, which appear in yellow-red colors.

Protein's radius of gyration. A biased estimate of a protein shape has consequences for studies on proteins folding. Stability and folding speed of a protein depend on both the estimated shape of the denatured state (unfolded state) and of the native state (folded state). One may study if compact initial states yield to faster folding. The protein compactness is represented by the protein's Radius of Gyration, defined as: $R_g^2 = \frac{1}{N} \sum_{\text{non H atoms } i}^N (r_i - R_C)^2$, where N is the number of non-hydrogen atoms, r_i , R_C are resp. the coordinates of atoms and centers. Error on atoms coordinates give a bias on the estimate of the Radius of Gyration:

$$(13) \quad \mathbb{B}(R_g^2) = \sigma^2 \frac{3(N-1)}{N} = \bar{R}_g^2 \frac{(N-1)}{N} \frac{3}{\text{SNR}^2},$$

where we also express this bias with respect to an adaptation of the signal-noise-ratio introduced in Section 3: $\text{SNR}^2 = \frac{R_g^2}{\sigma^2}$.

The radius of protein HJSJ (85 residues) is known to be around 10 Å. The error on R_g^2 is of 0.3% with an average error of positions on the atoms of 0.3Å. It is 8.6% for an error of 1.7Å. The error will be greater if one considers binding sites at the periphery of the proteins rather than the whole protein. Indeed sites' size is smaller and they have less atoms.

One could think about doing clustering on radii of Gyration using the K-means algorithm on shapes. The index D of Section 3 is:

$$(14) \quad D = \frac{R_1^{\sigma^2} - R_2^{\sigma^2}}{\sigma} = \frac{R_1^2 - R_2^2}{\sigma} + 3\sigma \left(\frac{N_1 - 1}{N_1} - \frac{N_2 - 1}{N_2} \right).$$

Clustering on radii of gyration may lead to a misleading indicator. D indicates that the clustering performs better than it actually does.

False positive probability in protein's motif detection. The relation between a protein's shape and function is linked to its motifs, which define the supersecondary structure. Motifs have biological properties: for example the helixturnhelix motif [9] is responsible for the binding of DNA within several prokaryotic proteins. Automatic motif detection is another challenge in the study of protein shapes. We investigate the impact of bias on the false positive probability estimation in motif detection.

Let us consider a set $\{P_i\}_{i=1}^n$ of proteins each with N_i atoms. One is interested in the motifs of k atoms that can be detected in the protein's set, where $k < N$. We define σ that represents an allowed error zone. The number of detected motifs increases if: (i) one decreases k , or (ii) one increases σ , or (iii) increases n . Thus how many detected motifs actually come from chance, with respect to the parameters k, σ, n ? The false positives probability indicates when one detects truth and when one detects noise. The usual estimate of the false positive probability is $P = \frac{\mathcal{V}_0}{\mathcal{V}_l}$. Here \mathcal{V}_0 is the volume of the error zone allowed. \mathcal{V}_l is the total volume of the protein [35], thus the a ball of radius the Radius of Gyration. Thus \mathcal{V}_l may be biased and overestimated. The probability of false positive is underestimated.

We consider the example of [37]. One tries to find motifs between the tryptophan repressor of Escherichia coli (PDB code 2WRP) and the CRO protein of phage 434 (PDB code 2CRO). These two proteins are known to share the helix-turn-helix motif. The radius of Gyration of 2WRP is $R_g = 20\text{Å}$, the total volume is: $\mathcal{V}_l = \frac{4}{3}\pi R_g^3 \simeq 33510\text{Å}^3$. We assume an error zone that takes the form of a diagonal covariance matrix with standard deviations $\sigma = 0.35\text{Å}$. We get the error zone volume: $\mathcal{V}_0 = \chi^{\frac{3}{2}} \frac{4}{3}\pi \sigma^3 = 4.06\text{Å}^3$, where $\xi^2 = 8$ comes from a convention about how much error is allowed: the covariance of the error within the error volume shall be less than ξ^2 , see [37] for details. The estimation of the false positive probability is: $P = 1.2 \times 10^{-4}$. We find that P is underestimated by 0.27% using the expression of the Radius of Gyration's bias.

4.4. Brain template in Neuroimaging. We apply the rule of thumb of Section 3 to determine when the bias needs a correction in the computation of a brain template from medical images. There are numerous technical difficulties for this application. First, M and G are now infinite dimensional. Then, the Lie group action is not necessarily isometric. Thus, it is

clear that the results of the theorems do not apply directly. Nevertheless, this subsection still allows us to gain intuition about how this paper may impact the field of neuroimaging.

In neuroimaging, a template is an image representing a reference anatomy. Computing the template is often the first step in medical image processing. Then, the subjects' anatomical shapes may be characterized by their spatial deformations *from the template*. These deformations may serve for (i) a statistical analysis of the subject shapes, or (ii) for automated segmentation by mapping the template's segmented regions into the subject spaces. In both cases, if the template is not centered among the population, i.e. if it is biased, then the analyzes and conclusions could be biased. We are interested in highlighting the variables that control the template's bias.

The framework of Large Deformation Diffeomorphic Metric Mapping (LDDMM) [41] embeds the template estimation in our geometric setting. The Lie group of diffeomorphisms acts on the space of images as follows:

$$(15) \quad \rho : \text{Diff}(\Omega) \times L_2(\Omega) \rightarrow L_2(\Omega), \quad (\phi, I) \mapsto \phi \cdot I = I \circ \phi^{-1}.$$

The isotropy group of I writes: $G_I = \{\phi \in \text{Diff}(\Omega) | I \circ \phi^{-1} = I\}$. Its Lie algebra \mathfrak{g}_I consists of the infinitesimal transformations whose vector fields are parallel to the level sets of I : $\mathfrak{g}_I = \{v | \forall x \in \Omega, \nabla I(x)^T \cdot v(x) = 0\}$. The orbit of I is: $O_I = \{I' \in L_2(\Omega) | \exists \phi \in \text{Diff}(\Omega) \text{ s.t. } I' \circ \phi^{-1} = I\}$.

The "shape space" is by definition the space of orbits. Two images that are diffeomorphic deformations of one another are in the same orbit. They correspond to the same point in the shape space. Topology of an image is defined as the image's properties that are invariant by diffeomorphisms. Consequently, the shape space is the space of the images topology, represented by the topology of their level sets. We get a stratification of the shape space when we gather the orbits by orbit type. A stratum is more singular than another, if it has higher orbit type, i.e. larger isotropy group.

The manifold M has an infinite stratification. One changes stratum every time there is a change in the topology of an image's level sets. Singular strata are connected to simpler topology. "Principal" strata are connected to more complicated topology. Indeed, the simpler the topology of the level sets is, the higher is the "symmetry" of the image. Thus the larger is its isotropy group. Note that strata with smaller isotropy group (more detailed topology) do not represent "singularities" from the point of view of a given image and do not influence the bias. In fact, such strata are at distance 0: an infinitesimal local change in intensity can create a maximum or minimum, thus complexifying the topology.

Using the rule-of-thumb of Section 3, the template's bias depends on its distance d to the next singularity, at the scale of σ the intersubjects variability. The template is biased in the regions where the difference in intensity between maxima and minima is of the same amplitude as the variability. The template may converge to pure noise in these regions.

Conclusion. We introduced tools of statistics on manifolds to study the properties of template's shape estimation in Medical imaging and Computer vision. We have shown asymptotic bias by considering the shape space's geometry. The bias comes from the external curvature of the template's orbit at the scale of the noise on the data. This provides a geometric interpretation for the bias observed in [2, 3]. We investigated the case of several templates and

the performance K-mean algorithms on shapes: clusters are less well separated because of each centroid's bias. The variables controlling the bias are: (i) the distance in shape space from the template to a singular shape and (ii) the noise's scale. This gives a rule-of-thumb for determining when the bias is important and needs correction. We proposed two procedures for correcting the bias: an iterative bootstrap and a nested bootstrap. These procedures can be applied to any type of shape data: landmarks, curves, images, etc. They also provide a way to compute the external curvature of an orbit.

Our results are exemplified on simulated and real data. Many studies use the template's shape estimation algorithm in Molecular Biology, Medical Imaging or Computer vision. Their estimations are necessarily biased. But these studies often belong to a regime where the bias is not important (less than 0.1%). For example, the bias is important in landmark shapes analyses when the landmarks' noise is comparable to the template shape's size. Studies are rarely in this regime. We have considered shapes belonging to infinite dimensional shape spaces. Our results do not apply to the infinite dimensional case. We have used them to gain intuition about it. The bias might be more important in infinite dimensions and needs a correction as we have suggested.

REFERENCES

1. D. ALEKSEEVSKY, A. KRIEGL, M. LOSIK, AND P. W. MICHOR, *The Riemannian geometry of orbit spaces. the metric, geodesics, and integrable systems*, Publ. Math. Debrecen, 62 (2003).
2. S. ALLASSONNIÈRE, Y. AMIT, AND A. TROUVÉ, *Towards a coherent statistical framework for dense deformable template estimation*, Journal of the Royal Statistical Society, 69 (2007), pp. 3–29.
3. S. ALLASSONNIÈRE, L. DEVILLIERS, AND X. PENNEC, *Estimating the template in the total space with the Fréchet mean on quotient spaces may have a bias.*, in Proceedings of the fifth international workshop on Mathematical Foundations of Computational Anatomy (MFCA'15), 2015, pp. 131–142.
4. S. ALLASSONNIÈRE, L. DEVILLIERS, AND X. PENNEC, *Fréchet means top and quotient space may not be consistent. (personal communication)*, (2016).
5. S. ALLASSONNIÈRE AND E. KUHN, *Convergent stochastic expectation maximization algorithm with efficient sampling in high dimension. application to deformable template model estimation*, Computational Statistics & Data Analysis, 91 (2015), pp. 4 – 19.
6. H. M. BERMAN, J. WESTBROOK, Z. FENG, G. GILLILAND, T. N. BHAT, H. WEISSIG, I. N. SHINDYALOV, AND P. E. BOURNE, *The protein data bank*, Nucleic Acids Res, 28 (2000), pp. 235–242.
7. J. BIGOT AND B. CHARLIER, *On the consistency of Fréchet means in deformable models for curve and image analysis*, Electronic Journal of Statistics, (2011), pp. 1054–1089.
8. J. BIGOT AND S. GADAT, *A deconvolution approach to estimation of a common shape in a shifted curves model*, Ann. Statist., 38 (2010), pp. 2422–2464.
9. R. G. BRENNAN AND B. W. MATTHEWS, *The helix-turn-helix dna binding motif.*, Journal of Biological Chemistry, 264 (1989), pp. 1903–6.
10. L. BREWIN, *Riemann normal coordinate expansions using cadabra*, Classical and Quantum Gravity, 26 (2009), p. 175017.
11. H. DARMANTÉ, B. BUGNAS, R. B. D. DOMPSURE, L. BARRESI, N. MIOLANE, X. PENNEC, F. DE PERETTI, AND N. BRONSARD, *Analyse biométrique de l'anneau pelvien en 3 dimensions à propos de 100 scanners*, Revue de Chirurgie Orthopédique et Traumatologique, 100 (2014), pp. S241–.
12. I. DRYDEN AND K. MARDIA, *Statistical shape analysis*, John Wiley & Sons, New York, 1998.
13. J. DU, I. L. DRYDEN, AND X. HUANG, *Size and shape analysis of error-prone shape data*, Journal of the American Statistical Association, 110 (2015), pp. 368–379.
14. B. EFRON, *Bootstrap methods: Another look at the jackknife*, The Annals of Statistics, 7 (1979), pp. 1–26.

15. J. EHLERS, F. A. E. PIRANI, AND A. SCHILD, *Republication of: The geometry of free fall and light propagation*, General Relativity and Gravitation, 44 (2012), pp. 1587–1609, doi:10.1007/s10714-012-1353-4, <http://dx.doi.org/10.1007/s10714-012-1353-4>.
16. A. M. T. ELEWA, *Morphometrics for Nonmorphometricians*, Springer, 2012.
17. M. ÉMERY AND G. MOKOBODZKI, *Sur le barycentre d'une probabilité dans une variété*, Séminaire de probabilités de Strasbourg, 25 (1991), pp. 220–233.
18. A. EVANS, A. JANKE, D. COLLINS, AND S. BAILLET, *Brain templates and atlases*, Neuroimage, 62(2) (2012), pp. 911–922.
19. C. GOODALL, *Procrustes Methods in the Statistical Analysis of Shape*, Journal of the Royal Statistical Society. Series B (Methodological), 53 (1991), pp. 285–339.
20. J. C. GOWER AND G. B. DIJKSTERHUIS, *Procrustes problems*, vol. 30 of Oxford Statistical Science Series, Oxford University Press, Oxford, UK, January 2004, <http://oro.open.ac.uk/2736/>.
21. S. HUCKEMANN, T. HOTZ, AND A. MUNK, *Intrinsic shape analysis: Geodesic principal component analysis for riemannian manifolds modulo lie group actions.*, Statistica Sinica, 20 (2010), pp. 1–100.
22. S. JOSHI, D. KAZISKA, A. SRIVASTAVA, AND W. MIO, *Riemannian structures on shape spaces: A framework for statistical inferences*, in Statistics and Analysis of Shapes, Birkhaeuser Boston, 2006, pp. 313–333.
23. D. G. KENDALL, *The diffusion of shape*, Advances in applied probability, 9 (1977), pp. 428–430.
24. D. G. KENDALL, *Shape manifolds, Procrustean metrics, and complex projective spaces*, Bulletin of the London Mathematical Society, 16 (1984), pp. 81–121.
25. S. A. KURTEK, A. SRIVASTAVA, AND W. WU, *Signal estimation under random time-warps and non-linear signal alignment*, in Advances in Neural Information Processing Systems 24, J. Shawe-taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger, eds., 2011, pp. 675–683.
26. H. LE, *On the consistency of procrustean mean shapes*, Advances in Applied Probability, 30 (1998), pp. 53–63.
27. H. LE AND D. G. KENDALL, *The riemannian structure of euclidean shape spaces: A novel environment for statistics*, The Annals of Statistics, 21 (1993), pp. 1225–1271.
28. S. LELE, *Euclidean distance matrix analysis (EDMA): estimation of mean form and mean form difference*, Mathematical Geology, 25 (1993), pp. 573–602.
29. J.-Y. LI, E. ENGLUND, J. HOLTON, D. SOULET, P. HAGELL, A. LEES, T. LASHLEY, N. QUINN, S. REHNCRONA, A. BJORKLUND, H. WIDNER, T. REVESZ, O. LINDVALL, AND P. BRUNDIN, *Lewy bodies in grafted neurons in subjects with parkinson's disease suggest host-to-graft disease propagation*, Nature, 14 (2008).
30. M. LORENZI, N. AYACHE, G. B. FRISONI, AND X. PENNEC, *Mapping the effects of $A\beta_{1-42}$ levels on the longitudinal changes in healthy aging: hierarchical modeling based on stationary velocity fields*, in Proceedings of Medical Image Computing and Computer Assisted Intervention (MICCAI), vol. 6892 of LNCS, Springer, 2011, pp. 663–670.
31. A. LYCHAK AND G. THORBERGSSON, *Curvature explosion in quotients and applications*, J. Differential Geom., 85 (2010), pp. 117–140.
32. L. MARCO AND X. PENNEC, *Parallel Transport with Pole Ladder: Application to Deformations of Time Series of Images*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 68–75.
33. N. MIOLANE AND X. PENNEC, *Biased estimators on quotient spaces*, Proceedings of the 2nd international of Geometric Science of Information (GSI'2015), (2015).
34. V. PATRANGENARU AND L. ELLINGSON, *Nonparametric Statistics on Manifolds and Their Applications to Object Data Analysis*, Taylor & Francis, 2015, <https://books.google.fr/books?id=z6mJSQAACAAJ>.
35. X. PENNEC, *Toward a generic framework for recognition based on uncertain geometric features*, Videre: Journal of Computer Vision Research, 1 (1998), pp. 58–87.
36. X. PENNEC, *Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements*, Journal of Mathematical Imaging and Vision, 25 (2006), pp. 127–154.
37. X. PENNEC AND N. AYACHE, *A geometric algorithm to find small but highly similar 3d substructures in proteins.*, Bioinformatics, 14 (1998), pp. 516–522.
38. M. POSTNIKOV, *Riemannian Geometry*, Encyclopaedia of Mathem. Sciences, Springer, 2001.
39. RSTUDIO, INC, *Easy web applications in R.*, 2013, <http://www.rstudio.com/shiny/>.
40. I. J. TICKLE, R. A. LASKOWSKI, AND D. S. MOSS, *Error Estimates of Protein Structure Coordinates*

and Deviations from Standard Geometry by Full-Matrix Refinement of γB - and $\beta B2$ -Crystallin, Acta Crystallographica Section D, 54 (1998), pp. 243–252.

41. L. YOUNES, *Shapes and Diffeomorphisms*, Applied Mathematical Sciences, Springer London, Limited, 2012.

Appendix A. Notation. We summarize in this appendix the main elements of the proofs of theorems 1 and 2.

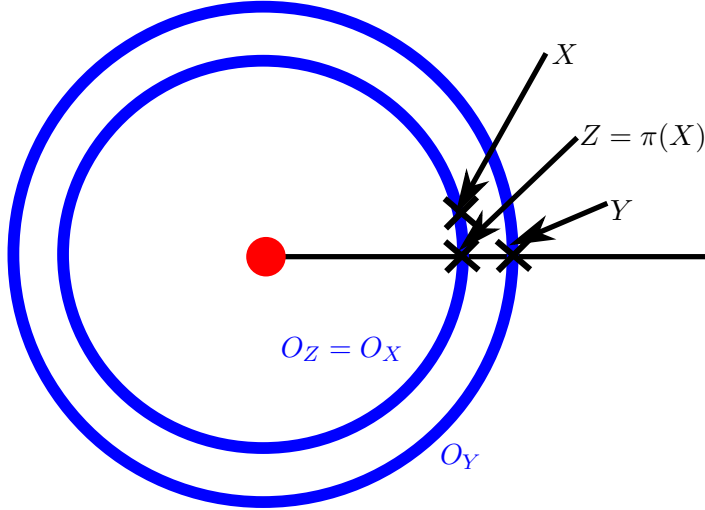


Figure 23. Summary of the notations used in the proofs.

Points (independent of coordinates systems). Points are denoted with upper cases. Y is the template shape Y and \hat{Y} is its estimate. Here X is a point in M . We consider that X belongs to a principal orbit. This will have no impact on the integration because the set of principal orbits is dense in M . We write $Z = \pi(X)$ the projection of X in the shape space. Figure 23 shows the elements Y , X , O_Z , $\pi(X)$. We denote m the dimension of M , p the dimension of the principal orbits and q the dimension of the quotient space.

Normal coordinate systems. We often use Normal Coordinate Systems (NCS) to express the coordinates of tangent vectors. We refer to [38] for theoretical developments about the NCS and to [10] for Taylor expansions of differential geometric tensors in a NCS.

For example, we may consider a NCS centered at the point Y , with respect to the Riemannian metric of M . This NCS is valid on an open neighborhood of Y , that is start-shaped domain around Y . Moreover, we assume that M is geodesically complete : thus, this domain is equal to the whole manifold M , with the exception of the cut locus which is of null measure [38]. The Riemannian logarithm of a point X in the NCS at Y is denoted: \overrightarrow{YX} .

Appendix B. Preliminaries.

B.1. Truncated Gaussian moments in a Euclidean space. We give preliminary computations of Gaussian moments in a m -dimensional vector space \mathbb{R}^m , using the curved notation \mathcal{M} . We refer to the (unnormalized) moment of order k of the m -dimensional Gaussian of

covariance $\sigma^2 A$ as:

$$\mathcal{M}^{i_1 \dots i_k}(\sigma^2 A) = \int_{\mathbb{R}^m} X^{i_1} \dots X^{i_k} \exp\left(-\frac{X^T A^{-1} X}{2\sigma^2}\right) dX$$

The order 0, 2 and 4 are:

$$(16) \quad \begin{aligned} \mathcal{M}^0(\sigma^2.A) &= \sigma^m \sqrt{(2\pi)^m} \sqrt{\det(A)} \\ \mathcal{M}^{ab}(\sigma^2.A) &= \sigma^{m+2} \sqrt{(2\pi)^m} \sqrt{\det(A)}. A^{ab} \\ \mathcal{M}^{abcd}(\sigma^2.A) &= \sigma^{m+4} \sqrt{(2\pi)^m} \sqrt{\det(A)}. (A^{ab} A^{cd} + A^{ac} A^{bd} + A^{ad} A^{bc}) \\ \mathcal{M}^{i_1 \dots i_k}(\sigma^2.A) &= \Theta\left(\sigma^{m+k}\right)^{i_1 \dots i_k} \quad \text{if } k \text{ even} \\ &= 0 \quad \text{if } k \text{ odd} \end{aligned}$$

where Θ denotes the proportionality to σ^{m+k} .

The *truncated (unnormalized) moment at radius r of the m -dimensional Gaussian of covariance $\sigma^2 A$* are defined as:

$$\mathcal{M}_r^{i_1 \dots i_k}(\sigma^2 A) = \int_{B_r} X^{i_1} \dots X^{i_k} \exp\left(-\frac{X^T A^{-1} X}{2\sigma^2}\right) dX$$

where the integration domain is now the m -dimensional ball B_r .

They write, with respect to the total moments:

$$(17) \quad \boxed{\mathcal{M}_r^{i_1 \dots i_k}(\sigma^2 A) = \mathcal{M}^{i_1 \dots i_k}(\sigma^2 A) + \epsilon(\sigma)}$$

where $\sigma \rightarrow \epsilon(\sigma)$ is a function that decreases exponentially for $\sigma \rightarrow 0$.

B.2. Isotropic Gaussian Moments on a Riemannian manifolds . We turn to computations of Gaussian moments in a m -dimensional Riemannian manifold M , using the notation \mathfrak{M} . We refer to the *(unnormalized) moment of order k of the m -dimensional isotropic Gaussian of covariance $\sigma^2 \mathbb{I}$* as:

$$\mathfrak{M}^{i_1 \dots i_k}(\sigma^2 \mathbb{I}) = \int_M \overrightarrow{YX}^{i_1} \dots \overrightarrow{YX}^{i_k} \exp\left(-\frac{d_M^2(X, Y)}{2\sigma^2}\right) dM(X)$$

and to the *truncated (unnormalized) moment at radius r of the m -dimensional isotropic Gaussian of covariance $\sigma^2 \mathbb{I}$* as:

$$\mathfrak{M}_r^{i_1 \dots i_k}(\sigma^2 \mathbb{I}) = \int_{B_r} \overrightarrow{YX}^{i_1} \dots \overrightarrow{YX}^{i_k} \exp\left(-\frac{d_M^2(X, Y)}{2\sigma^2}\right) d\overrightarrow{YX}$$

where the integration domain is now the m -dimensional geodesic ball B_r of radius r and centered at Y .

First, we consider the truncated moments:

$$(18) \quad \mathfrak{M}_r^{i_1 \dots i_k}(\sigma^2 \mathbb{I}) = \sigma^{k+m} \mathcal{M}^{i_1 \dots i_k}(\mathbb{I}) + \frac{\sigma^{k+m+2}}{6} \sum_{ab} R_{ab}(Y) \mathcal{M}^{abi_1 \dots i_k}(\mathbb{I}) + \mathcal{O}(\sigma^{m+k+4}) + \epsilon(\sigma)$$

They write, with respect to the non-truncated moments:

$$\mathfrak{M}^{i_1 \dots i_k}(\sigma^2 \mathbb{I}) = \mathfrak{M}_r^{i_1 \dots i_k}(\sigma^2 \mathbb{I}) + \int_{\mathcal{C}_{B_r}} X^{i_1} \dots X^{i_k} \exp\left(-\frac{d_M^2(Y, X)}{2\sigma^2}\right) dM(X)$$

The second term of the sum is negligible when $\sigma \rightarrow 0$.

$$(19) \quad \mathfrak{M}^{i_1 \dots i_k}(\sigma^2 \mathbb{I}) = \sigma^{k+m} \mathcal{M}^{i_1 \dots i_k}(\mathbb{I}) + \frac{\sigma^{k+m+2}}{6} \sum_{ab} R_{ab}(Y) \mathcal{M}^{abi_1 \dots i_k}(\mathbb{I}) + \mathcal{O}(\sigma^{m+k+4}) + \epsilon(\sigma).$$

where $\epsilon(\sigma)$ decreases exponentially when $\sigma \rightarrow 0$.

Appendix C. Proof of Theorem 1: Induced probability density on shapes.

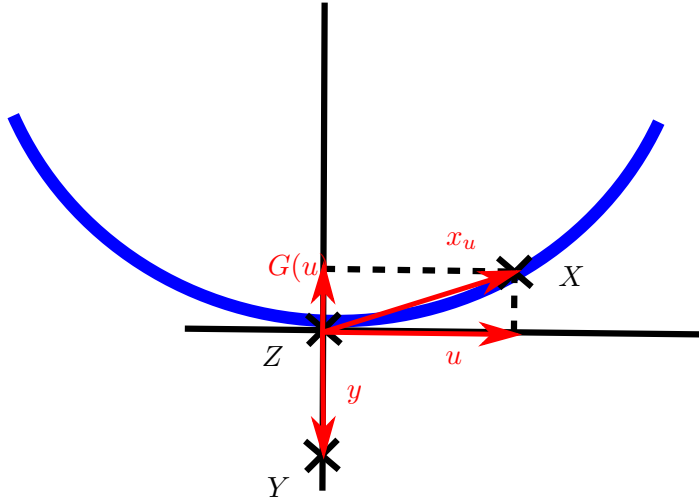


Figure 24. Summary of the notations used in the roof of Theorem 1.

The generative model implies the following Riemannian Gaussian distribution on the objects:

$$(20) \quad f(Z) = \frac{1}{C_M(\sigma)} \exp\left(-\frac{d_M^2(X, Y)}{2\sigma^2}\right),$$

where $C_M(\sigma)$ is the integration constant:

$$(21) \quad C_M(\sigma) = \int_M \exp\left(-\frac{d_M^2(X, Y)}{2\sigma^2}\right) dM(X)$$

We compute the induced probability distribution f on shapes by integrating the distribution on the orbit of X out of $f(Z)$:

$$f(Z) = \frac{1}{C_M(\sigma)} \int_{X \in O_Z} \exp\left(-\frac{d_M^2(Y, X)}{2\sigma^2}\right) dO_Z(X).$$

where $Z = \pi(X)$ is fixed in the following computations. We give a Taylor expansion of $f(Z)$ for $\sigma \rightarrow 0$ to the order σ^2 .

We start by dividing the integral:

$$f(Z) = \frac{1}{C_M(\sigma)} \cdot \left(\int_{O_Z \cap B_r} \exp\left(-\frac{d_M^2(Y, X)}{2\sigma^2}\right) dO_Z(X) + \int_{O_Z \cap \mathcal{C}_{B_r}} \exp\left(-\frac{d_M^2(Y, X)}{2\sigma^2}\right) dO_Z(X) \right) \blacksquare$$

In the parenthesis, we call the first integral $I_Z(\sigma)$ and the second term is an $\epsilon_Z(\sigma)$:

$$f(Z) = C_M(\sigma)^{-1} \cdot (I_Z(\sigma) + \epsilon_Z(\sigma))$$

We compute the Taylor expansions of $C_M(\sigma)^{-1}$ and $I_Z(\sigma)$ for $\sigma \rightarrow 0$ and show that $\epsilon_Z(\sigma) = o\left(\exp\left(-\frac{r^2}{2\sigma^2}\right)\right)$

C.1. Taylor expansion of $C_M(\sigma)^{-1}$. The integration constant $C_M(\sigma)$ is the moment of order 0 of the m -dimensional isotropic Gaussian of variance $\sigma^2 \mathbb{I}$ in the m -dimensional manifold M .

$$C_M(\sigma)^{-1} = (\sqrt{2\pi}\sigma)^{-m} \cdot \left(1 - \frac{\sigma^2}{6} R(Y) + O(\sigma^3) + \epsilon(\sigma)\right)$$

C.2. Taylor expansion of $I_Z(\sigma)$. Now we compute the integral $I_Z(\sigma)$ involved in the formula of $f(Z)$:

$$I_Z(\sigma) = \int_{O_Z \cap B_r} \exp\left(-\frac{d_M^2(Y, X)}{2\sigma^2}\right) dO(X)$$

We denote: $B_{r_Z} = B_{r_Z}$. We first perform the computations for any σ . We recall that r is fixed, and small enough in a sense made precise later.

C.2.1. Computing $d_M^2(Y, X)$. The first component of $I_Z(\sigma)$ is $d_M^2(Y, X)$. We express the Taylor expansion of d_M^2 for $X \in O_Z$ close to Z .

First, we parameterize the point X . The points X , $Z = \pi(X)$ and the orbit $O_Z \subset M$ are illustrated on Figure 23: the orbit $O = O_Z$ is the blue circle in $M = \mathbb{R}^2$ going through X and Z . The orbit O_Z can be seen in the tangent space $T_Z M$ through the Logarithm map at Z . For r small enough, i.e. for X close enough to Z , we can locally represent O_Z in $T_Z M$ as the graph of a smooth function G from $T_Z O$ to $N_Z O$, using the vector $u \in T_Z O_Z$ around $u = 0$:

$$\begin{aligned} I : T_Z O_Z &\mapsto T_Z M = T_Z O_Z \oplus N_Z O_Z \\ u &\mapsto x_u = (u, G(u)) \end{aligned}$$

The local graph $u \rightarrow G(u)$ is illustrated on Figure 23 and has the following Taylor expansion around $u = 0$ in the NCS at Z :

$$G(u)^a = \frac{1}{2}h_{bc}^a(Z)u^bu^c + G_3(Z)_{bcd}^au^bu^cu^d + G_4(Z)_{bcde}^au^bu^cu^du^e + O(\|u\|^5)$$

The 0-th and 1-th order derivatives of $u \rightarrow G(u)$ are zero because the graph goes through Z and is tangent at $T_Z O_Z$. The second order derivative is by definition the second fundamental form $h(Z)$ introduced in Section 2: $h(Z)$ represents the best quadratic approximation of the graph G . The third and fourth orders $G_3(Z)$ and $G_4(Z)$ are further refinements on the shape of the graph G around Z .

Second, we compute the Taylor expansion of $d_M^2(Y, X)$ with respect to u :

$$d_M^2(Y, X) = d_M^2(Y, Z) + u^a L_a(Z) + u^a u^b M_{ab}(Z) + u^a u^b u^c P_{abc}(Z) + u^a u^b u^c u^d S_{abcd}(Z) + O(\|u\|^5)$$

where:

$$(22) \quad \begin{aligned} L_a(Z) &= 0 \\ M_{ab}(Z) &= -y^c h_{ab}(Z)^c - \frac{1}{3} R_{acbd}(Z) y^c y^d + \frac{1}{12} \nabla_d R_{aebc}(Z) y^d y^e y^c \\ &\quad - \frac{1}{180} (44 R_{geaf}(Z) R_{gcbd}(Z) + 3 \nabla_{ef} R_{acbd}(Z)) y^e y^f y^c y^d \\ &\quad - \frac{1}{54} R_{hfa g}(Z) \nabla_c R_{hdbe}(Z) y^f y^g y^c y^d y^e \end{aligned}$$

and $P_{abc}(Z)$ and $S_{abcd}(Z)$ are tensors mixing the derivatives of the graph G and the derivatives of the Riemannian curvature R .

C.2.2. Computing $dO_Z(X)$. The second component of $I_Z(\sigma)$ is the measure of the orbit $dO_Z(X)$. We seek the Taylor expansion of the measure:

$$(23) \quad dO_Z(X) = \left(1 + T_c(Z)u^c - N_{cb}(Z)u^b u^c + O(\|u\|^3) \right) du$$

where:

$$(24) \quad \begin{aligned} T_c(Z) &= h_{ca}^a(Z) \\ N_{cb}(Z) &= \frac{1}{6} \text{Ric}_{cb}(Z) + \frac{1}{2} h_{ca}^a(Z) h_{ca}^a(Z) - \frac{1}{2} h_{cd}^a(Z) h_{ba}^d(Z) \end{aligned}$$

C.2.3. Gathering to compute $I_Z(\sigma)$. We plug the expressions of $d_M^2(Y, X)$ and $dO_Z(X)$, computed in the previous subsections, in the expression of $I_Z(\sigma)$ to get:

$$I_Z(\sigma) = \exp\left(-\frac{d_M^2(Y, Z)}{2\sigma^2}\right) (\sigma^p m_0(Z) + \sigma^{p+2} m_2(Z) + \Theta(\sigma^4) + \epsilon(\sigma))$$

where:

$$(25) \quad \begin{aligned} m_0(Z) &= \mathcal{M}(M)^0 = \sqrt{(2\pi)^p} \sqrt{\det((M_{ab}(Z))^{-1})} \\ m_2(Z) &= N_{cb}(Z) \mathcal{M}(M^{-1}(Z))^{bc} - \frac{1}{2} S_{abcd}(Z) \mathcal{M}(M^{-1}(Z))^{abcd} \end{aligned}$$

where M_{ab} and S_{abcd} are given in the previous subsections.

C.3. Upper bound on $\epsilon_Z(\sigma)$. Ultimately, we show that:

$$(26) \quad \epsilon_Z(\sigma) = o\left(\exp\left(-\frac{r^2}{2\sigma^2}\right)\right)$$

and ϵ_Z is exponentially decreasing when $\sigma \rightarrow 0$.

C.4. Final result: Taylor expansion of $f(Z)$. Replacing the terms in the expression of $f(Z)$ and get:

$$f(Z) = \frac{\exp\left(-\frac{d_M^2(Y,Z)}{2\sigma^2}\right)}{(\sqrt{2\pi}\sigma)^q} (F_0(Z) + \sigma^2 F_2(Z) + \Theta_Z(\sigma^4) + \epsilon(\sigma))$$

where:

$$(27) \quad \begin{aligned} F_0(Z) &= \sqrt{\det((M_{ab}(Z))^{-1})} \\ F_2(Z) &= -\frac{1}{6} \sqrt{\det((M_{ab}(Z))^{-1})} R(Z) + \frac{m_2(Z)}{(\sqrt{2\pi})^p} \end{aligned}$$

and we refer to the previous subsections for the formula of $M_{ab}(Z)$ and $m_2(Z)$.

Appendix D. Proof of Theorem 2: Bias on the template shape. We compute the bias $\text{Bias}(Y, \hat{Y})$ of \hat{Y} as an estimator of Y . In the following, we take a NCS at Y . In particular, the vector $\vec{Y\hat{Z}} = \text{Log}_Y Z$ has coordinates written z .

The expectation of the distribution f of shapes in Q is \hat{Y} by definition. The point \hat{Y} , expressed in a NCS at the template Y , gives $\text{Bias}(Y, \hat{Y})$, a tangent vector at $T_Y M$ that indicates how much one has to shoot to reach the estimator \hat{Y} :

$$(28) \quad \text{Bias}(Y, \hat{Y}) = \text{Log}_Y \hat{Y} = \int_Q \vec{Y\hat{Z}} f(Z) dQ(Z).$$

First, we take a ball of small radius r in Q and fix r . We split the integral:

$$\text{Bias}(Y, \hat{Y}) = \int_{B_r^Q} \vec{Y\hat{Z}} f(Z) dQ(Z) + \int_{C_{B_r^Q}} \vec{Y\hat{Z}} f(Z) dQ(Z)$$

By the result of the preliminaries, adapted to Q , the right part is a function $\sigma \rightarrow \epsilon(\sigma)$ that is exponentially decreasing for $\sigma \rightarrow 0$.

$$\text{Bias}(Y, \hat{Y}) = \int_{B_r^Q} \vec{Y\hat{Z}} f(Z) dQ(Z) + \epsilon(\sigma)$$

D.1. Using the result of Theorem 1. We plug the expression of the density f using Theorem 1, writing the Taylor expansions of the F 's terms around $z = 0$:

$$\begin{aligned} F_0(Z) &= F_{00}(Y) + F_{01d}(Y)z^d + \mathcal{O}(\|z\|^2) \\ F_2(Z) &= F_{20}(Y) + F_{21d}(Y)z^d + \mathcal{O}(\|z\|^2) \end{aligned}$$

to get:

$$\begin{aligned} \text{Bias}(Y, \hat{Y})^a &= F_{01d}(Y)\sigma^2\delta^{ad} + \epsilon(\sigma) \\ &= F_{01}^a(Y)\sigma^2 + \mathcal{O}(\sigma^4) + \epsilon(\sigma) \end{aligned}$$

D.2. Computation of $F_{01}^a(Y)$: Taylor expansion of $F_0(Z)$ in the coordinate z . The term $F_{01}^a(Y)$ is the first order coefficient in the Taylor expansion of $F_0(Z)$ around Y in the coordinate z . We compute its Taylor expansion first in y and then convert it into a Taylor expansion in z . We have $y = -z$ and we get:

$$F_{01}^a(Y) = -\frac{1}{2}H^a(Y)$$

D.3. Final result: Taylor expansion of $\text{Bias}(\hat{Y}, Y)$. Replacing $F_{01}^a(Y)$ by its value computed above:

$$\text{Bias}(Y, \hat{Y})^a = -\frac{1}{2}H^a(Y)\sigma^2 + \mathcal{O}(\sigma^4) + \epsilon(\sigma)$$

which is the final result.